PHASE I
LAUNCH COMMIT CRITERIA
PERFORMANCE TRENDING ANALYSIS

SRM&QA MISSION SERVICES

Contract No. NAS8-37812

Revision A

May 10, 1989

Prepared by

Advanced Technology, Inc.
555 Sparkman Drive, Suite 410
Huntsville, AL 35816

Prepared for

National Aeronautics and Space Administration
George C. Marshall Space Flight Center
Marshall Space Flight Center, AL 35812

TD-002-342-444-0002
SRM&QA

PHASE I
LAUNCH COMMIT CRITERIA
PERFORMANCE TRENDING ANALYSIS

SRM&QA MISSION SERVICES
Contract No. NAS8-37812 ✓

Revision A

May 10, 1989

Prepared by

Advanced Technology, Inc.
555 Sparkman Drive, Suite 410
Huntsville, AL 35816

Prepared for

National Aeronautics and Space Administration
George C. Marshall Space Flight Center
Marshall Space Flight Center, AL 35812

This report was prepared by Advanced Technology, Inc. (ATI) under contract NAS8-37812 (WBS 1.3.4.2) funding to assess quantitative methods and measures for monitoring Launch Commit Criteria (LCC) performance measurement trends. The specific Technical Directive (TD-002-342-444-0000) directed that a statistical performance trending analysis pilot study, utilizing four selected SRB, SRM, ET, and SSME Shuttle measurement types from the five missions prior to 51-L, be processed and compared to STS-26 mission data.

To accommodate this technical guidance, under the statistical uncertainties of small sample theory, representative LCC measurement types were selected by Marshall Space Flight Center (MSFC) personnel from the Reliability and Maintainability (R&M) Office, with contractor support from Boeing Aerospace Operations (BAO), Calspan Corporation, and ATI. The guideline criteria utilized in this selection process placed high priority on the statistical stability of the candidate measurements, in opposition to measurement criticality, to insure that the selected measurements were representative of the complete measurement spectrum. As a result, the following measurement types were selected:

1) Solid Rocket Booster (SRB) Auxiliary Power Unit (APU) turbine speed (rpm)
2) External Tank (ET) Liquid Hydrogen (LH_2) ullage pressure (psi)
3) Space Shuttle Main Engine (SSME) Low Pressure Fuel Turbopump (LPFTP) discharge temperature (^{o}R)
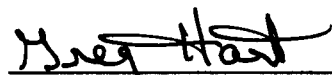4) Range Safety Switch (RSS) safe and arm device (event).

These measurement types comprise three analog and one discrete class of measurements which range from highly volatile oscillations to a simple on-off signal.

Once the raw data coordinates were obtained, each set of measurements within a measurement type was processed to obtain statistical confidence bounds and mean data profiles for each of the selected measurement types. STS-26 measurements were compared to the statistical data base profiles to verify the statistical capability of assessing:

1) occurrence of data trend anomalies and
2) abnormal time-varying operational conditions associated with data amplitude and phase shifts.

Prepared By: Approved By:

_____   _____
Greg Hart   Gordon Hoskins
R&M Engineering   R&M Director

_____   _____
Mary Bestor   Dr. Pat R. Odom
R&M Engineering   Project Manager

_____
Roger Cross
R&M Engineering

TABLE OF CONTENTS

TABLE OF CONTENTS (Concluded)

LIST OF FIGURES AND TABLES

LIST OF FIGURES AND TABLES (Concluded)

1.0 INTRODUCTION

1.1 OBJECTIVE

To accommodate the tasking objectives of the Phase I Launch Commit Criteria (LCC) Trending Analysis (LCCTA) effort, an LCC performance trending plan was written which provided a sequential overview of the major steps involved in this prototype development. In summary, the development objectives of this proof-of-concept were oriented toward the:

- Methodology to be used for obtaining the required data to support development of suitable statistical confidence bounds and mean data profiles of four selected Shuttle LCC measurement types from the five missions prior to 51-L.

- Approach to be used in converting trending requirements and issues into relevant operational algorithms suitable for LCC trending evaluations.

1.2 SCOPE

To help direct the initial focus of the prototype effort, current trending techniques, requirements and issues were reviewed to identify the critical areas of concern. This review indicated that the primary focus of this effort should be directed toward the analytical formulation of an efficient "distribution-free" probability density function (pdf) and associated unbiased mean estimator for characterizing the time-varying statistical data base profiles for each selected measurement type.

This primary analytical focus resulted from the acknowledged realization that time-varying statistical data variations contained in historical successful mission measurements must be characterized by pdf's that are capable of representing the extreme variations of such data. The problem of selecting a particular form or type of pdf which best fits an empirical data sample has long been an open question in descriptive statistics.

In theory, this same problem carries over to the Gaussian or Laplacian normal probability distribution. That is, no physical distribution can rigorously be a normal universe, as this distribution extends to infinity in both directions.

A similar statement applies to any universe whose distribution extends to infinity in either direction. For this reason, it is often convenient to approximate the distribution from a random data sample of the parent population. Nevertheless, the a priori requirement of specifying a distribution (form) remains an unanswered question. As a result, a highly intensive literature search and mathematical investigation was undertaken in the Phase I effort to identify a satisfactory probability distribution form. The final selection was a special case of the Gram-Charlier Series

(GCS) approximation which utilizes the standardized normal distribution as a generating function.

This basic series is based on the restrictive premise that the parent distribution has the following properties:

● Moments of all order exist.

● Derivatives of any required order exist with appropriate continuity.

● There exists high order contact at the extremities of the distribution. (See Appendix B, Section B.2.)

Given these properties and the added restriction that the parent population is defined on an infinite interval, an approximating distribution can be generated by the sum of a system of independent skewed frequency distributions of the Gram-Charlier type. In principle, this series expresses the required parent population in terms of the derivatives of the standardized normal distribution with zero mean and unit variance. It also allows for the recursive series coefficient formulation while maintaining explicit asymptotic conformity to the central limit theorem.

The notion of extreme measurement variation is also amplified in the development of the statistical data base mean profiles. The critical concern is that historical measurement data often reflect displacements which could be related to design changes or time sequence shifts in the operational aspects of the elements being monitored. In statistical terms, these variations appear as outliers and should not be given full recognition unless these effects are persistent from mission to mission. To accommodate this relative statistical mean stability problem, an unbiased mean estimator was formulated which de-emphasizes the relative time-varying worth of rapidly changing historical measurement observations.

2.0 ANALYSIS APPROACH

2.1 CONCEPT OVERVIEW

To achieve a computationally efficient proof-of-concept assessment of the quantitative methods for monitoring LCC performance, a statistical confidence bound approach was conceptualized which comprises both real-time and preprocessing components. The conceptual overview of these processing components is illustrated in Figure 2.1-1.

2.1.1   Data Segmentation

From the preprocessing component perspective, the historical data coordinates, (t, x_t), of each measurement within a specified LCC measurement type were assigned to one of the following data

2

START
INPUT: segmented empirical data coordinates, $[t, x_t]$, for each of the successful missions and measurement types under consideration

SELECT: data segment and measurement type

COMPUTE: constituent segment frequencies, W_j, for each selected measurement type segment

COMPUTE: scaling coefficients A_j, and B_j, A_0, for each selected category for $j=1,2,...,M_k$

DETERMINE: unbiased segment mean estimator

$$F_k(t) = A_0 + \sum_{j=1}^{M_k} [A_j \cos W_j t + B_j \sin W_j t]$$

IDENTIFY: statistical confidence bounds for the specified confidence level α

Historical Mission Data Base

Trend Report Data Base

INPUT: candidate segment report

ASSESS: candidate report comparison with statistical bounds and flag violations

i^{th} Data Segment k^{th} Measurement Type

x_t

t

Segment

Segmented Flight Reports

Unbiased Segment Mean

Maximum Redline

Upper Bound

Unbiased Segment Mean

Lower Bound

Confidence Interval at Time t

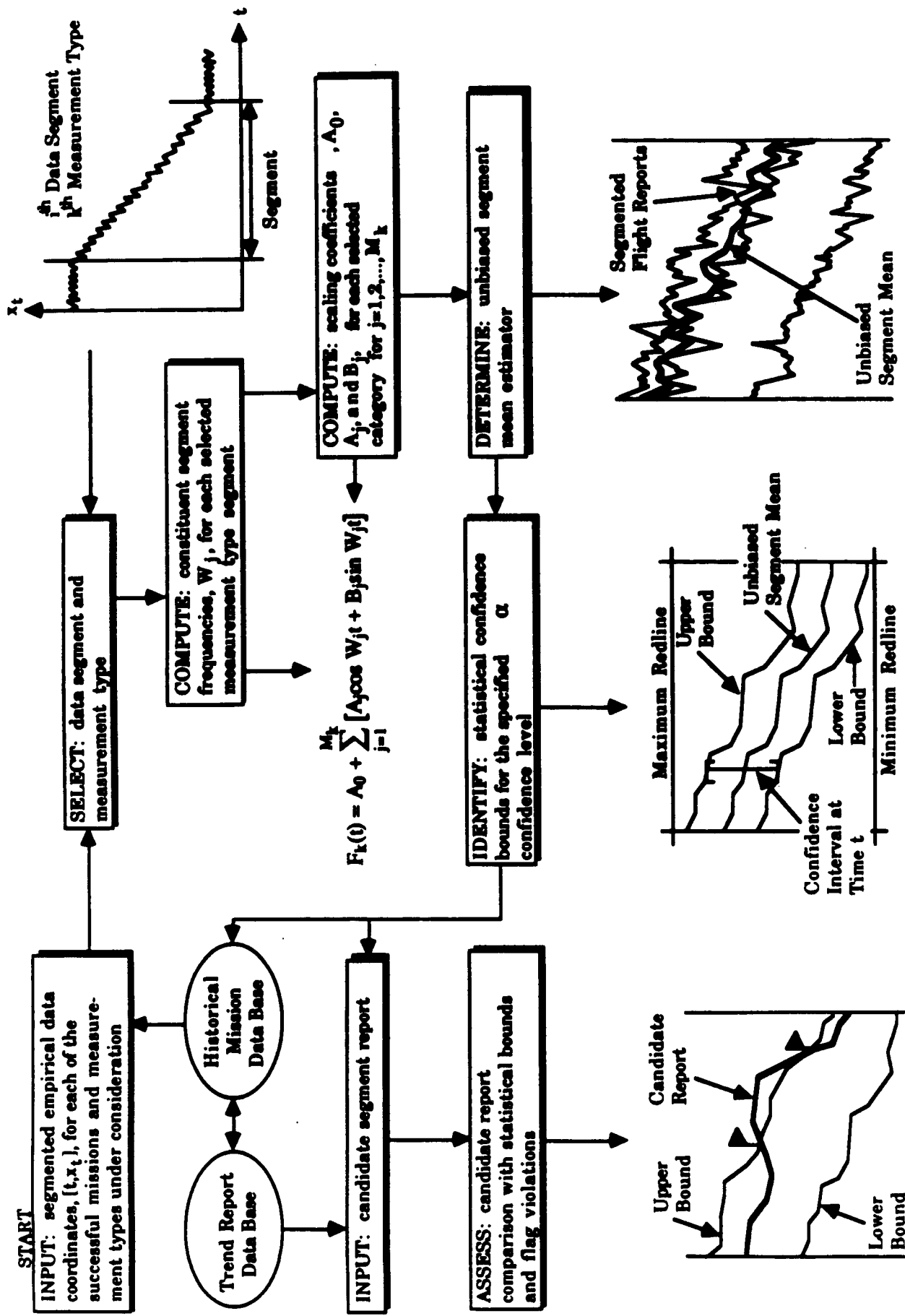Minimum Redline

Upper Bound

Candidate Report

Lower Bound

Figure 2.1-1 Conceptual Overview

3

segment types:

1) mandatory segment
2) constant segment
3) dropout segment
4) spontaneous active segment.

Under this segmentation criteria, a mandatory data segment is defined as a critical measurement time interval in which safe operational amplitude variation limits (redlines) are identified. By comparison, a constant data segment is defined as a time interval which has constant amplitude over its entire domain of definition. In juxtaposition, a data dropout segment is defined as a time-interval in which the data coordinates have been lost because of data dropouts resulting from such causes as telecommunications, computer transfer, etc. Finally, a spontaneous active data segment is any other time interval not meeting the above criteria.

Having segmented the appropriate measurements of a given type from previous successful missions, these individual data segment intervals can be fit by a proprietary data sampling and smoothing algorithm. The underlying structure of this innovative algorithm is designed to characterize oscillatory periodic signals defined by discrete data points at near real-time data acquisition rates.

2.1.2 Data Segment Fitting Algorithm

To provide this capability the governing periodicity equation,

$$F(W) = Cos[mWh] - C_1 Cos[(m-1)Wh] - C_2 Cos[(m-2)Wh] - \ldots - C_{m-2} Cos[2Wh] - C_{m-1} Cos[Wh] - C_m/2 = 0$$

for each measurement segment is solved for the constituent (harmonic) signal frequencies, W_j, for j=1, 2, ..., m. Here, h is the equidistant data spacing of the respective measurement coordinates and m denotes the number of distinct periods embedded in the imposed data coordinates. The C_1, C_2, ..., C_m coefficients of this governing periodicity equation are determined such that the linear combination of functional terms

$$f(t) = A_0 + \sum_{j=1}^{m} [A_j Cos(W_j t) + B_j Sin(W_j t)]$$

provides the necessary accuracy of the required empirical data fit.

To accelerate the processing requirements of this imposed solution for large data sets, an orthogonal vector projection scheme is utilized with recursive back substitution to compute the A_j and B_j scaling coefficients. This process allows for all computations completed in one step to be used in the next. In addition, it provides for controlled accuracy assessment during the computations. If at any point the length of one of

4

the projected vectors becomes zero (or near zero), the particular data sample associated with this vector is linearly dependent when compared to the previous data samples. Thus, this inherent recognition property provides the capability to perceive the limits of sufficient data, i.e., number of data points. In application, this process is continued over a finite data set to remove the possibility of a spurious data sample. If this trend continues, the projective vector space basis is satisfied and the remaining coefficients become zero, which provides a critical error deflation feature.

2.1.3   Data Preprocessing Component

The next preprocessing component phase statistically determines the unbiased segment mean profile. The formulation associated with this statistical determination utilizes an unbiased estimator of minimum variance to de-emphasize the relative time-varying worth of rapidly changing historical measurement observations. A detailed explanation of this formulation is provided in Section 2.2 for those interested in the mathematical details of this unbiased estimator.

2.1.4   Data Confidence Bound Determination

Having determined the unbiased segment mean profiles, a confidence level, α, is specified to reflect the percentage of measurements within the defined limits of the statistical confidence bound profiles. This evaluation utilizes the "distribution-free" GCS approximation to characterize the sampled time-varying parent (measurement) pdf. From this pdf, the time-varying upper and lower confidence bounds are computed. These bounds are obtained by setting the integral of the pdf equal to $(1-\frac{\alpha}{2})$ and $\alpha/2$, respectively, and then solving for the associated upper and lower limits of integration. The mathematical details of this formulation process are presented in Section 2.3.

If this bounding process is repeated for

$$t_j = t_I + (j-1)\Delta t \text{ for } j=1, 2, 3, \ldots$$

equidistant coordinates, the amplitude values of the upper and lower confidence bounds are defined. Continuous functional representations of these bounding profiles (i.e., of the above noted form) are then stored in the historical mission data base for future real-time component processing. A similar continuous representation of the time-varying mean profile is also stored in the historical mission data base.

It is this integrated data base profile storage organization that provides the essential interface between the real-time and preprocessing component environments. Figure 2.1-1 illustrates this elemental processing feature by the two-way arrows between the historical mission and trend report data bases.

5

2.1.5   Real-Time Data Processing

In real-time data processing, the segmented candidate measurement to be trended is superimposed on its associated confidence level profiles obtained from the historical mission data base and any out-of-bound conditions are identified.

2.1.6   Confidence Bound Application

In application, this confidence level concept provides an additional data assessment capability. Specifically, various confidence levels can be specified to identify aggregated confidence bands or regions which characterize the relative time-varying measurement amplitudes from their mean measurement dispersion. This basic confidence bound flexibility is based on the individual confidence bound proposition that $100(1-\alpha)$ percent of the parent measurement population is expected to fall within a specified $(1-\alpha)$ confidence interval.

2.2   UNBIASED MEAN ESTIMATOR

The problem of determining an unbiased mean estimator in the presence of statistical outliers has eluded many investigators. For example, the iteration-of-means procedure introduced in the Phase I LCC performance trending plan reflects this problem. That is, the order in which the x_t's were taken affected the final iteration-of-means value. In addition, the procedure outlined in the above referenced plan imposed the restriction that the measurements within a segmented time interval had to be statistically independent due to original plan formulation.

2.2.1   De-Emphasized Relative Measurement Worth

Given the above restrictive provisions, other techniques were explored for de-emphasizing the relative time-varying worth of rapidly changing historical measurement observations. A weighted mean of minimum variance approach was utilized to accommodate this de-emphasis.

To account for time-varying measurement amplitudes around a particular time, t_j, the influence of previous and later time variations were considered in this approach. Figure 2.2-1 illustrates the statistical methodology used to assimilate these time-varying measurement effects.

6
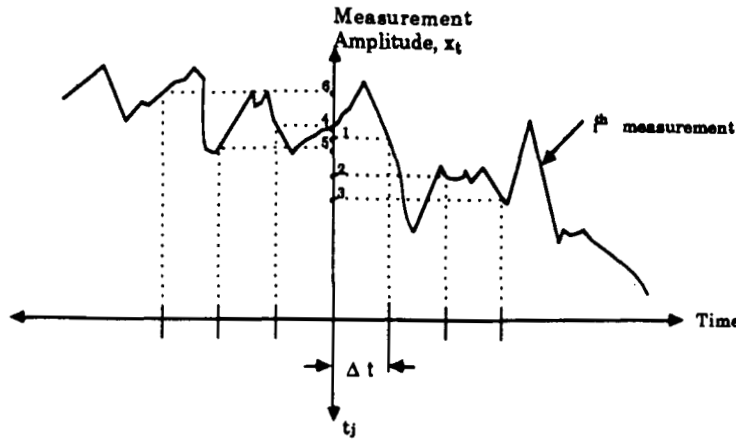
Figure 2.2-1   Assimilation of Time-Varying Measurement Effects

The amplitude of a measurement is sampled at equally spaced intervals on either side of the time of interest, t_j. The resulting amplitude images are then projected on the amplitude time line. At this point, the projected images are considered representative statistical samples of the referenced (i^{th}) measurement variation in the vicinity (neighborhood) of t_j.

2.2.2   Weighted Mean of Minimum Variance

Imposition of the constraint that the average image point for each measurement has the same expectation, combined with the condition that the linear combination of averages is unbiased and of minimum variance, completely defines the required mean estimator.

If one imposes the condition that each image point average is defined as:

$$\bar{x}_k = \frac{1}{n} \sum_{j=1}^{n} x_{kj},$$

where x_{kj} is the jth image point of the kth measurement and n is the total number of image points used to characterize the measurement variation, the above noted linear combination of $\bar{x}_k$'s becomes

$$\hat{x} = \sum_{i=1}^{k} a_i \bar{x}_i,$$

where $\hat{x}$ is the unbiased mean estimator and $\bar{x}_i$ represents the mean of the ith measurement for k distinct measurements. For this expression to be unbiased, the a_i weights must be constrained so that

$$\sum_{i=1}^{k} a_i = 1.$$

That is, for $\hat{x}$ to be unbiased and for each image point average to have the same expectation, $E[\bar{x}] = E[\bar{x}_i]$,

$$E[\hat{x}] = E[\sum_{i=1}^{k} a_i \bar{x}_i]$$

7

$$= \sum_{i=1}^{k} a_i E[\bar{x}_i]$$

$$= E[\bar{x}_i] \sum_{i=1}^{k} a_i$$

$$= E[\bar{x}].$$

Similarly, the variance of this unbiased mean estimator, $\hat{x}$, is

$$VAR[\hat{x}] = VAR \left[\sum_{i=1}^{k} a_i \bar{x}_i \right]$$

$$= \sum_{i=1}^{k} a_i^2 \; VAR[\bar{x}_i]$$

$$= \sum_{i=1}^{k} a_i^2 \sigma_i^2$$

where σ_i^2 equals the variance of the ith measurement image point sample variation, i.e.,

$$\sigma_i^2 = \frac{1}{n-1} \sum_{j=1}^{n} (x_{ij} - \bar{x}_i)^2.$$

To determine the values of a_i which make the variance of $\hat{x}$ a minimum, subject to the condition of unbiasedness, note that

$$a_k + \sum_{i=1}^{k-1} a_i = \sum_{i=1}^{k} a_i = 1.$$

Therefore,

$$a_k = 1 - \sum_{i=1}^{k-1} a_i.$$

Thus, the variance of $\hat{x}$ can be written in terms of $a_1, a_2, \ldots, a_{k-1}$:

$$VAR[\hat{x}] = \sum_{i=1}^{k-1} a_i^2 \sigma_i^2 + a_k^2 \sigma_k^2$$

$$= \sum_{i=1}^{k-1} a_i^2 \sigma_i^2 + \left[1 - \sum_{i=1}^{k-1} a_i \right]^2 \sigma_k^2$$

$$= \sum_{i=1}^{k-1} a_i^2 \sigma_i^2 + \left[1 - 2\sum_{i=1}^{k-1} a_i + \left(\sum_{i=1}^{k-1} a_i \right)^2 \right] \sigma_k^2.$$

To determine the values of a_i which make this a minimum, one differentiates with respect to a_j and equates the result to zero:

$$\frac{\partial VAR[\hat{x}]}{\partial a_j} = 2[a_j \sigma_j^2 - (1 - \sum_{i=1}^{k-1} a_i) \sigma_k^2]$$

$$= 2a_j \sigma_j^2 - 2a_k \sigma_k^2$$

$$= 0.$$

8

Hence, $a_j = \dfrac{a_k \sigma_k^2}{\sigma_j^2}$ for j=1, 2, ..., k-1. However, this relationship is obviously true for j=k. Consequently, one can sum over j=1, 2, ..., k to get

$$\sum_{j=1}^{k} a_j = a_k \sigma_k^2 \sum_{j=1}^{k} \frac{1}{\sigma_j^2}.$$

But,

$$\sum_{j=1}^{k} a_j = 1 \text{ so that}$$

$$a_k = \frac{1}{\sigma_k^2 \sum\limits_{j=1}^{k} \dfrac{1}{\sigma_j^2}}.$$

Substituting this result into the above expression,

$$a_j = \frac{1}{\sigma_j^2 \sum\limits_{i=1}^{k} \dfrac{1}{\sigma_i^2}}.$$

Upon noting that $\sum\limits_{i=1}^{k} \dfrac{1}{\sigma_i^2}$ is constant for all values of j, the associated $\bar{x}_j$ weights are inversely related to the variance of the jth measurement variation about its mean. In other words, measurements with relatively large variances are assigned accompanying small weights under this unbiased mean estimator.

2.3 CONFIDENCE BOUND FORMULATION

This section summarizes the algebraic derivation of the upper and lower confidence bound formulation used in this Phase I effort. In theory, this upper bound derivation involves setting the definite pdf integral equal to $1-\frac{\alpha}{2}$ and solving for the upper limit. Since the pdf is defined in terms of the GCS,

$$1 - \frac{\alpha}{2} = \int_{-\infty}^{z_u} \left[\sum_{n=0}^{\infty} (-1)^n \frac{C_n}{n!} \Psi^{(n)}(z) \right] dz$$

$$= \int_{-\infty}^{z_u} \Psi(z) dz + \sum_{n=2}^{\infty} \left[(-1)^n \frac{C_n}{n!} \int_{-\infty}^{z_u} \Psi^{(n)}(z) dz \right]$$

which results from $C_0=1$ and $C_1=0$. See Appendix A, Section A.4, for an overview of the mathematical details involved.

2.3.1 Algebraic Rational Approximation

Utilizing Hasting's rational polynomial approximation (see Reference 1) to evaluate the standard normal probability,

$$\int_{-\infty}^{z_u} \Psi(z) dz = \frac{1}{\sqrt{2\pi}} \int_{-\infty}^{z_u} e^{\frac{-z^2}{2}} dz$$

9

becomes:

$$P(z_u) = \int_{-\infty}^{z_u} \Psi(z)\,dz = 1 - \Psi(z_u)\sum_{j=1}^{5} b_j r^j + \epsilon(z_u)$$

where

$$\Psi(z_u) = \frac{1}{\sqrt{2\pi}}\, e^{\frac{-z_u^2}{2}}$$

$$r = \frac{1}{1+pz_u}$$

$$|\epsilon(z_u)| < 7.5 \times 10^{-8}$$

$$
\begin{aligned}
p &= 0.2316419 \\
b_1 &= 0.319381530 \\
b_2 &= -0.356563782 \\
b_3 &= 1.781477937 \\
b_4 &= -1.821255978 \\
b_5 &= 1.330274429
\end{aligned}
$$

Taking this approximation in conjunction with the relations

$$\int_{-\infty}^{z_u} \Psi^{(n)}(z)\,dz = \Psi^{(n-1)}(z_u) \quad \text{and} \quad \Psi^{(n-1)}(z_u) = (-1)^{n-1}\Psi(z_u)H_{n-1}(z_u)$$

one obtains the following algebraic expression for the upper confidence bound:

$$g(z_u) = \Psi(z_u) \sum_{n=1}^{5} b_n \left[\frac{1}{1+pz_u} \right]^n - \sum_{n=2}^{\infty} \frac{C_n}{n!} \Psi(z_u)H_{n-1}(z_u) - \frac{\alpha}{2} = 0.$$

Having shown in Appendix A, Section A.2, that the product of the Gaussian error curve and $(n-1)^{th}$ order Hermite polynomial, $H_{n-1}(z_u)$, can be written in the recursive form

$$\Psi(z_u)H_{m+1}(z_u) = G_{m+1}(z_u)$$

$$= \frac{1}{m+1}\left[z_u G_m(z_u) - G_{m-1}(z_u) \right]$$

with initial starting conditions

$$G_0(z_u) = \Psi(z_u) \quad \text{and} \quad G_1(z_u) = z_u G_0(z_u).$$

It follows that the optimal upper confidence bound is a root of the equation:

$$g(z_u) = \Psi(z_u) \sum_{n=1}^{5} b_n \left[\frac{1}{1+pz_u} \right]^n - \sum_{n=2}^{\infty} \frac{C_n}{n!} G_{n-1}(z_u) - \frac{\alpha}{2} = 0.$$

2.3.2   Numerical Root Evaluation

Before considering the numerical procedure used to extract this root, it is interesting to note that experience has shown that only four to six GCS coefficients are generally required to provide adequate series convergence. The exception is extremely skewed distributions. Nevertheless, the infinite limit of the right-most summation can be replaced by a finite integer in actual application. Consequently,

$$g(z_u) = \Psi(z_u) \sum_{n=1}^{5} b_n \left[\frac{1}{1+pz_u} \right]^n - \sum_{n=2}^{N} \frac{C_n}{n!} G_{n-1} (z_u) - \frac{\alpha}{2} = 0$$

where N equals the number of coefficients (terms) utilized in the GCS characterization.

To invoke computational efficiency, an iteration procedure originally introduced by Müller in 1956 (see Reference 2) was used to consecutively compute approximations to the desired root. This determination was obtained by substitution of parabolic for linear interpolation in the derivation of the "reguli falsi" or method of false position (see Reference 3) to produce the iterative equation

$$z_{i+1} = z_i - \frac{2c_i}{b_i \pm \sqrt{b_i^2 - 4a_i c_i}}, \text{ for } i = 2, 3, \ldots .$$

Here the sign preceding the radical is chosen to make the absolute value of the denominator as large as possible. This sign convention results in z_{i+1} being selected as the closest root to z_i of the approximating parabolic interpolation

$$p(z_u) = a_i(z_u - z_i)^2 + b_i(z_u - z_i) + c_i = 0.$$

Since this formulation requires that $p(z_u)$ pass through the points $(z_{i-2}, g(z_{i-2}))$, $(z_{i-1}, g(z_{i-1}))$ and $(z_i, g(z_i))$, the coefficients a_i, b_i, and c_i at the i+1 iteration step are determined from the conditions:

$$c_i = g(z_i)$$

$$b_i = \frac{(z_{i-2} - z_i)^2[g(z_{i-1}) - g(z_i)] - (z_{i-1} - z_i)^2[g(z_{i-2}) - g(z_i)]}{(z_{i-2} - z_i)(z_{i-1} - z_i)(z_{i-2} - z_{i-1})}$$

$$a_i = \frac{(z_{i-1} - z_i)[g(z_{i-2}) - g(z_i)] - (z_{i-2} - z_i)[g(z_{i-1}) - g(z_i)]}{(z_{i-2} - z_i)(z_{i-1} - z_i)(z_{i-2} - z_{i-1})}.$$

To satisfy these conditions, three approximating roots are needed at each iteration step. This imposes the question of how to select the initial three approximating guesses. Under the assumption that the LCC measurement distributions are constructively (approximately) normally distributed, at each selected evaluation time t_j, an initial approximation for z_i can be obtained by realizing that

$$\frac{\alpha}{2} = \int_{z_i}^{\infty} \Psi(z) dz = \frac{1}{\sqrt{2\pi}} \int_{z_i}^{\infty} e^{\frac{-z^2}{2}} dz$$

and then using a rational approximation to determine z_i. That is,

$$z_i = S - \frac{2.30753 + 0.270615 \, S}{1.0 + 0.992295 \, S + 0.04481 \, S^2} + \epsilon \left[\frac{\alpha}{2} \right]$$

11

where $0 < \frac{\alpha}{2} \le 0.5$, $S = \sqrt{\ln\left[\frac{4}{\alpha^2}\right]}$ and $\left| \epsilon\left[\frac{\alpha}{2}\right] \right| < 3 \times 10^{-3}$. Under this course of reasoning, the initial values of z_2 and z_0 are defined by adding and subtracting an arbitrarily small constant from z_1.

Once the a_2, b_2, and c_2 coefficients are determined from these initial starting conditions, the z_3 iterate is computed and the procedure is reinitialized using z_1, z_2, and z_3 in place of z_0, z_1, and z_2 to determine the next approximation z_4. At this point the procedure continues until a root is obtained within a prescribed a priori tolerance, i.e., $|z_{i+1} - z_i| \le$ tol. Table 2.3-1 outlines the underlying algorithmic steps involved in this procedure.

2.3.3   Summary Remarks

As noted in step 4 of Table 2.3-1, this procedure involves the evaluation of the radical $\sqrt{b^2 - 4ac}$ at each iteration step. Consequently, the procedure has the capability to determine complex roots, should they exist. Furthermore, the provision for supporting the determination of the lower confidence bound, z_1, is provided in this same algorithmic framework. To obtain z_1 one need only substitute $2-\alpha$ for α and impose the constraint that $z_u = -z_u$.

The rationale behind these conditions is reflected by the provision that

$$\frac{\alpha}{2} = \int_{-\infty}^{-z_1} \left[\sum_{n=0}^{N} (-1)^n \frac{C_n}{n!} \Psi^{(n)}(z) \right] dz.$$

In summary, an analysis of the convergence properties of this procedure for the selected Phase I LCC measurements showed that, on the average, only 14 iterations were required to obtain the upper and lower confidence bounds within a prescribed accuracy of 0.00032 measurement units. By comparison, the extreme iteration counts ranged from 1 to 47. Because of the assumed normal rational approximation utilized in the initialization process, normally distributed measurements require only one iteration, while highly skewed measurement distributions require substantially more.

12

TABLE 2.3-1. MÜLLER'S ALGORITHMIC STRUCTURE

Given the confidence level α; upper confidence bound equation, $g(t)=0$; maximum number of iterations, N_0; and required upper bound (root) tolerance, tol:

Step 1   Establish initial starting iterate guesses:

$$S = \sqrt{\ln\left(\frac{4}{\alpha^2}\right)};$$

$$t_1 = S - \frac{2.30753 + 0.270615\ S}{1.0 + 0.992295\ S + 0.04481\ S^2};$$

$$t_0 = t_1 - \delta_0; \quad \{\delta_0 = \text{arbitrary small constant}\}$$

$$t_2 = t_1 + \delta_0;$$

Step 2   Set $h_1 = \delta_0$;

$h_2 = \delta_0$;

$\delta_1 = [g(t_1) - g(t_0)]/h_1$;

$\delta_2 = [g(t_2) - g(t_1)]/h_2$;

$d = [\delta_2 - \delta_1]/[h_2 + h_1]$;

$i = 2$

Step 3   While $i \le N_0$, do Steps 4-8.

Step 4 $b = \delta_2 + h_2 d$;

$D = (b^2 - 4g(t_2)d)^{\frac{1}{2}}$ {May require complex arithmetic.}

Step 5   If $|b-D| < |b+D|$ then set $E = b+d$, else set $E = b-d$.

Step 6   Set $h = -2g(t_2)/E$;

$p = t_2 + h$.

Step 7   If $|h| < $ tol then
OUTPUT (p)   {Procedure successful}
STOP.

Step 8   Set $t_0 = t_1$;   {Prepare for next iteration}

$t_1 = t_2$;

$t_2 = p$;

$h_1 = t_1 - t_0$;

$h_2 = t_2 - t_1$;

$\delta_1 = [g(t_1) - g(t_0)]/h_1$;

$\delta_2 = [g(t_2) - g(t_1)]/h_2$;

$d = [\delta_2 - \delta_1]/[h_2 + h_1]$;

$i = i+1$

Step 9   OUTPUT - Method failed after N_0 iterations.
{Procedure unsuccessful}
STOP.

2.4   INTEGRATED BOUND FORMULATION

Having outlined the statistical assumptions and computational features of the unbiased mean estimator and confidence bound formulations, this discussion addresses the integrated implementation aspects of the integrated bound formulation process. To facilitate an understanding of this three step process, a diagrammatic overview of these steps is provided in Figure 2.4-1.
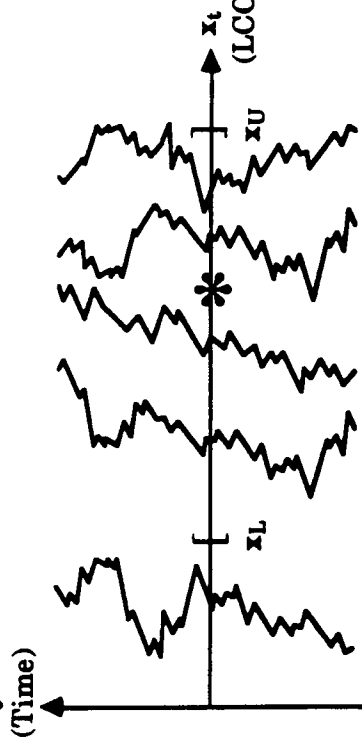
The pertinent implementation aspects of the first step recapitulates the necessity of considering an unbiased mean estimator to account for statistical measurement outliers related to design changes or time sequence shifts in the operational measurement displacements. Once the mean estimator is established for a referenced evaluation time, t_j, the aggregated spectrum of assimilated time-varying image points can be utilized to generate a histogram of the measurement variations relative to the referenced observation time of interest. This histogram of aggregated measurement variations is depicted in Figure 2.4-1 as the second step of the process. The theoretical aspects of this histogram development will be discussed in Section 2.4.1 along with Sheppard's correction for mid-point class frequencies.

Before the above noted mid-point class frequencies can be used in the third step of this process to compute the GCS coefficients and upper and lower confidence bounds, they must be transformed into z-coordinate moments by the standard z transformation of Figure 2.4-1. The theoretical development of this requirement follows from the general properties of semi-invariants under a linear transformation. These generalized properties are discussed in Section 2.4.2 for those interested in the mathematical details involved.

2.4.1   Histogram Development

Histogram development for characterizing the shape of a distribution from discrete sample data observations suffers from the unavoidable restriction that such representations are likely to exhibit large errors, unless the number of sample observations is large. This is reflected by the actual construction of a histogram. The range of the observation data is divided into arbitrary bins or class intervals, and the number or proportion of observations falling in each interval is assigned to the mid-point of the interval. As a result, small data samples often have class intervals which do not contain any observations unless the class interval length is increased. At the other extreme, the selected class interval could be so large that all observations fall in a single class interval.

14

1. Raw LCC measurements with unbiased mean estimator, $*$

$$x_U = z_U \sigma_x + \hat{x}$$
$$x_L = z_L \sigma_x + \hat{x}$$

3. Gram-Charlier pdf with upper and lower bounds

Incorporates change in both origin and scale

$$z = \frac{x_t - \hat{x}}{\sigma_x}$$

2. Generated raw data histogram frequency

Figure 2.4-1 Representative Bound Formulation Process

2.4.1.1   Application Implications

Regarded in this light, the class interval length selected for each aggregated spectrum of assimilated time-varying image points affects the smoothness of the resulting histogram. If the class interval is too small, the histogram is ragged; if it is too large, the histogram is over-smoothed and information is obscured.   Although   important,   the   influence   of   histogram smoothness is somewhat minimized by the constructive formulation of the current application.   Fundamentally, the least squares minimization imposed on the form of the GCS (see Appendix A, Section A.3), combined with the constructive addition of assimilated time-varying image points, diminishes this effect.

In recent years, this smoothing effect has received increased attention.   For example, Rudemo (see Reference 4) has developed automatic smoothing estimation methods from the probability density kernel, by letting W(x) be a nonnegative, symmetric weight function, centered at zero and integrating to one.   Under this method, W(x) can be the standard normal density function so that

$$W_h(x) = \frac{1}{h} W\left(\frac{x}{h}\right)$$

is a rescaled version of W where h is the class interval bandwidth of the estimating function.   As h approaches zero, W_h becomes more concentrated and peaked about zero.   As h approaches infinity, W_h becomes more spread out and flatter.   Under the assumption that W(x) is a standard normal density, $W_h(x)$ is a normal density with standard deviation h.

In general terms, if x_1, x_2, ..., x_n is a sample drawn from an unknown pdf, f, an estimate of f is given by

$$f_h(x) = \frac{1}{n} \sum_{i=1}^{n} W_h(x-x_i).$$

Such a function is called a kernel probability density estimate, since it consists of the superposition of "hills" centered over the sampled observations.   In the case where W(x) is the standard normal density, $W_h(x-x_i)$ is the normal density with mean x_i and standard deviation h.

In general, the bandwidth parameter, h, controls the estimating function smoothness and corresponds to the class interval length of the histogram.   Theoretically, an optimum bandwidth value can generally be established under this approach by defining an a priori smoothness criteria, providing the number of sample observations are large enough to support the imposed criteria.

16

This histogram smoothing methodology, among others, was investigated during the Phase I effort. The conclusion was that all such approaches provided little if any benefit over the straight GCS minimization property augmented by assimilated time-varying image point construct. In fact, certain combinations of circumstances provided significantly worse histogram estimates.

2.4.1.2   Constructive Histogram Development

Under the realization that the basic serial GCS gives the frequency function in the form

$$\Phi(z) = \sum_{n=0}^{N} a_n \Psi^{(n)}(z),$$

where the various coefficients are expressed as moments or semi-invariants, the concern over histogram development methodology has a secondary effect. The important consideration is that moments are exclusively expressed as definite integrals which are often difficult to determine in extremely skewed distributions. Moreover, unless the observations are numerous, it is generally impossible to compute moments of higher order than six. In the current application, this results from the large errors arising from the random sampling and measurement variations about the unbiased mean estimator. As a rule of thumb, it is generally useless to compute moments from a histogram of higher order than four to six when the number of individual observations is less than 400 to 500. This rule is once again confirmed by the assimilated time-varying image point observations; only the lower limit must be extended to approximately 460 for this application.

Failure to abide by this rule results in exhibiting a somewhat "poor fit" to the originating observation data and often gives rise to negative frequencies at the extreme tails of the associated GCS distribution. From a purely practical consideration, this last objection has little effect on the upper and lower confidence bounds, because the originating observations at the distribution extremities are very few in number. As a result, the constructive histogram development approach implemented in the Phase I effort divided the observation data into class intervals centered over the unbiased mean estimator in such a way that at least every interval contained one or more observations.

2.4.1.3   Adjusted Method of Moments

In basic statistics, the kth moment of a particular frequency is defined as the product of the frequency times the kth power of the distance about which the moment is required. To utilize this

17

definition to evaluate the moments of a relative histogram about
a selected point, one is forced to assume that the various
histogram frequencies are concentrated at the mid-point of their
respective class intervals. In addition, the problem of
determining a reference point about which to take moments that
produces a pre-assigned first moment

$$M_1 = \frac{\sum\limits_{j=1}^{N} S_j f_j}{\sum\limits_{j=1}^{N} f_j}$$

remains an open issue, since the S_j distances are unknown
variables. However, the denominator of the above moment
expression equals 1.0, given that the f_j's are relative
histogram frequencies. The solution to this problem can be
further motivated by letting the distance between A, the point
about which the moments are known, and B, the point about which
they are required, be d. Then, if the distance of any histogram
frequency, f_j, from A is X_j, and from B is x_j, $x_j = X_j - d$ and $x_j^k = (X_j - d)^k$.

To illustrate how the above distance relationships can be
used to evaluate d, consider the hypothetical histogram data of
Table 2.4-2. In this illustrative data sample of 1,000
measurements, the class interval length is 5 measurement units
and the imposed unbiased mean estimator equals 0.480. The
numbers -4, -3, -2, ... in column (3) are the respective
distances from the class interval mid-point values from d, in
terms of the unit of grouping. By forcing the condition that
the unbiased mean estimator, $\hat{x}$, must equal M_1, the expression

$$S_j = \frac{X_j - d}{h}$$

is imposed where h is the assigned histogram class interval
length. Therefore,

$$\hat{x} = \sum_{j=1}^{N} S_j f_j = \sum_{j=1}^{N} \left[\frac{X_j - d}{h} \right] f_j$$

$$= \frac{1}{h} \left[\left(\sum_{j=1}^{N} X_j f_j \right) - d \right]$$

from the fact that the f_j's are relative frequencies, i.e.,

$$\sum_{j=1}^{N} f_j = 1.$$

Consequently,

$$d = \sum_{j=1}^{N} X_j f_j - h\hat{x}.$$

Therefore, the adjusted class interval distance is

$$d = 79.400 - (5)(0.480) = 77.000$$

for this illustrative example. As a result, the totals of
columns (4) to (7) are the respective first, second, third, and
fourth discrete moments about $\hat{x}$.

18

Table 2.4-2 Adjusted Method of Moments

Adjusted Method of Moments

Class Interval Index j	Class Interval Mid-Point Value X_j (1)	Relative Mid-Point Frequencies f_j (2)	Adjusted Class Interval Distance $S=\dfrac{X_j-77}{5}$ (3)	$M_1=f_jS_j$ (4)	$M_2=f_jS_j^2$ (5)	$M_3=f_jS_j^3$ (6)	$M_4=f_jS_j^4$ (7)
1	57	0.029	-4	-0.116	0.464	-1.856	7.424
2	62	0.023	-3	-0.690	0.207	-0.621	1.863
3	67	0.081	-2	-0.162	0.324	-0.648	1.296
4	72	0.151	-1	-0.151	0.151	-0.151	0.151
5	77	0.192	0	0.000	0.000	0.000	0.000
6	82	0.239	1	0.239	0.239	0.239	0.239
7	87	0.157	2	0.314	0.628	1.256	2.512
8	92	0.093	3	0.279	0.837	2.511	7.533
9	97	0.029	4	0.116	0.464	1.856	7.424
10	102	0.006	5	0.030	0.150	0.750	3.750
Totals				0.480	3.464	3.336	32.192

Total number of measurements = 1000

Recalling the assertion of Section 2.4.1.2 that moments are exclusively expressed as definite integrals, a correction for treating histogram observations as concentrated frequencies at the class interval mid-points will be discussed in the remainder of this section. The theoretical aspects of this undertaking are focused on the desire to calculate the magnitude of discrete moment adjustments so that the true moment magnitudes can be assessed.

Given that the values of the discrete moments, M_k, are obtained from finite sums and not as definite integrals, they are subject to certain adjustments if one wishes to express them as continuous moments. The magnitudes of these adjustments can be computed from well-known formulas from the theory of numerical quadrature (see Appendix B) if the frequency function and its derivatives vanish for $x = -\infty$ and $x = +\infty$. The English mathematician, Sheppard, has developed the following correction formulas for the transition from discrete (M_k) to continuous (μ_k) moments:

$$\mu_0 = M_0$$

$$\mu_1 = M_1$$

$$\mu_2 = M_2 - \frac{h^2}{12} M_0$$

$$\mu_3 = M_3 - \frac{h^2}{4} M_1$$

$$\mu_4 = M_4 - \frac{h^2}{2} M_2 + \frac{7h^4}{240} M_0$$

$$\mu_5 = M_5 - \frac{5h^2}{6} M_3 + \frac{7h^4}{48} M_1$$

$$\vdots$$

$$\mu_k = \sum_{n=0}^{k} \binom{k}{n} (2^{1-n} - 1) B_n M_{k-n} h^n$$

where h equals the assigned class interval length and the B_n coefficients are the Bernoulli numbers generated by the recursive form

$$B_{k-1} = - \frac{1 + \binom{k}{1} B_1 + \binom{k}{2} B_2 + \ldots + \binom{k}{k-2} B_{k-2}}{\binom{k}{k-1}}$$

with starting value $B_0 = 1$.

These correction formulas can also be used to assess the effects of roundoff errors. For example, the third correction formula shows that a mean error of $\frac{h}{2}$ in x affects M_2 only as $\frac{h^2}{12}$. In summary, Sheppard's corrections emphasize the fact that the method of moments works with curve areas instead of curve ordinates.

2.4.2 Semi-Invariant Linear Transformation Properties

The evaluation of the GCS coefficients (using the semi-invariant relationships outlined in Appendix A, Section A.4) defined by the moments in the x-coordinate (measurement) system must be transformed into z coordinates by the transformation,

$$z = \frac{x - \hat{x}}{\sigma}.$$

This implies both a change of origin and scale. Specifically, the origin is transformed to z=0 and the scale is altered by the scaling factor $\frac{1}{\sigma}$, where σ is the parent measurement population standard deviation estimate.

Denoting the operator form of the x-coordinate system semi-invariants by $\lambda_1(x)$, $\lambda_2(x)$, $\lambda_3(x)$, ... with transformed origin $\lambda_1(z)$, the general linear transformation z=ax+b becomes:

$$\lambda_1(z) = \lambda_1(ax+b)$$
$$= a\lambda_1(x)+b$$
$$\lambda_2(z) = \lambda_2(ax+b)$$
$$= a^2\lambda_2(x)$$

Letting $\lambda_1(x) = \hat{x}$ and $\lambda_2(x) = \sigma^2$,

$$\lambda_1(z) = a\hat{x}+b$$
$$\lambda_2(z) = a^2\sigma^2.$$

Since the coordinate system of reference (z system) must have a zero mean and unit variance, $\lambda_1(z)=0$ and $\lambda_2(z)=1$ so that

$$a\hat{x}+b = 0$$
$$a\sigma = 1.$$

From this relationship $a = \frac{1}{\sigma}$ and $b = -\frac{\hat{x}}{\sigma}$, which corresponds to the above transformation form.

Moreover, $\lambda_j(z) = \frac{1}{\sigma^j} \lambda_j(x)$ for all values of j greater than 2. Consequently, the following simple rules can be used to epitomize the general semi-invariant computations:

- Set $\lambda_1(x) = \hat{x}$ the unbiased mean estimator.

- Compute $\lambda_j(x)$ for all values of j equal to or greater than 2. The numerical values of these paramenters divided by σ^j, for j=2, 3, 4, ... are the semi-invariants of the histogram frequency in the z system of the GCS.

21

3.0 DEVELOPMENT FINDINGS

3.1 DATA TRANSFER COMPLICATIONS

To accommodate the data transfer requirements of the Phase I effort, the individual raw data coordinates for each measurement had to be identified and transferred to 1.2 megabytes, 5 1/4 inch diskettes. Rockwell International, Inc. was tasked to collect and transfer this mission data to ATI. This data included the referenced missions 51-I, 51-J, 61-A, 61-B, 61-C, and the STS-26 mission measurements used for proof-of-concept trending.

The raw data coordinates associated with these six missions constituted a data storage requirement of approximately 100 megabytes of data. Given that there were 17 individual measurements involved with the four different measurement types per mission, each measurement represented an average data storage of 980 kilobytes.

During the verification of data readability, it was determined that there existed time gaps in the acquired data. This problem was attributed to data dropouts resulting from such sources as telecommunications, computer transfer, etc. To overcome this problem, numerous data corrections were required. To save time and effort, ATI was directed by Mr. Pizzano, chief of the MSFC Reliability and Maintainability Engineering Division, to consider only measurement data from T-600 seconds to T+600 seconds. As a result, all measurements were processed in this directed time interval during the data reduction phase of this effort.

### 3.2 SOFTWARE PARAMETER TUNING

During the early phase of prototype software development, it was recognized that individual confidence bound profiles would exhibit sufficiently different spatial behavior to warrant definition of specific processing parameters. To reconcile this acknowledgement, a parameter declaration option was included in the Parameter File Generation (PARMGEN) software module for individual segment processing parameter specification. In retrospect, the fortuitous inclusion of this option was confirmed and utilized for confidence bound segment profiles involving measurement variability.

To illustrate the rudimentary effects of confidence bound sensitivity, consider the tabular spectrum (Table 3.2-1) of image point generation parameters used to compute 97 percent confidence bound profiles for the highly variable SRB APU Turbine Speed LCC measurements. Assessment of the experienced confidence level bounds associated with various parameter combinations is summarized in Figure 3.2-1. Here the maximum and minimum deviations from the stipulated confidence level are the 50 and 26 image point cases shown graphically in Figures 3.2-10 and 3.2-6, respectively. In terms of percent confidence level variation, these same parameter combinations exhibit confidence bound differences of only 1.52 and

22

Table 3.2-1 Effects of Confidence Bound Sensitivity to Image Point Parameter Specification

Referenced Profile Shaping Figures	Delta Distance From t_j	Distance Between Adjacent t_j's	Assimilated* Image Pts. Per Measurement	Average Percent Measurements In Bounds
3.2-2	0.015625	0.03125	4	96.04
3.2-3	0.015625	0.06250	6	96.17
3.2-4	0.003906	0.03125	10	96.36
3.2-5	0.003906	0.03125	18	96.73
3.2-6	0.003906	0.03125	26	96.84
3.2-7	0.015625	0.06250	34	97.45
3.2-8	0.003906	0.03125	42	98.29
3.2-9	0.015625	0.03125	46	98.33
3.2-10	0.015625	0.06250	50	98.52
3.2-11	0.015625	0.03125	58	95.89
3.2-12	0.007813	0.06250	74	95.72
Overall Average				96.94

* For 20 Measurement Sample Size, i.e. 4 Measurements Per Mission Over 5 Missions.
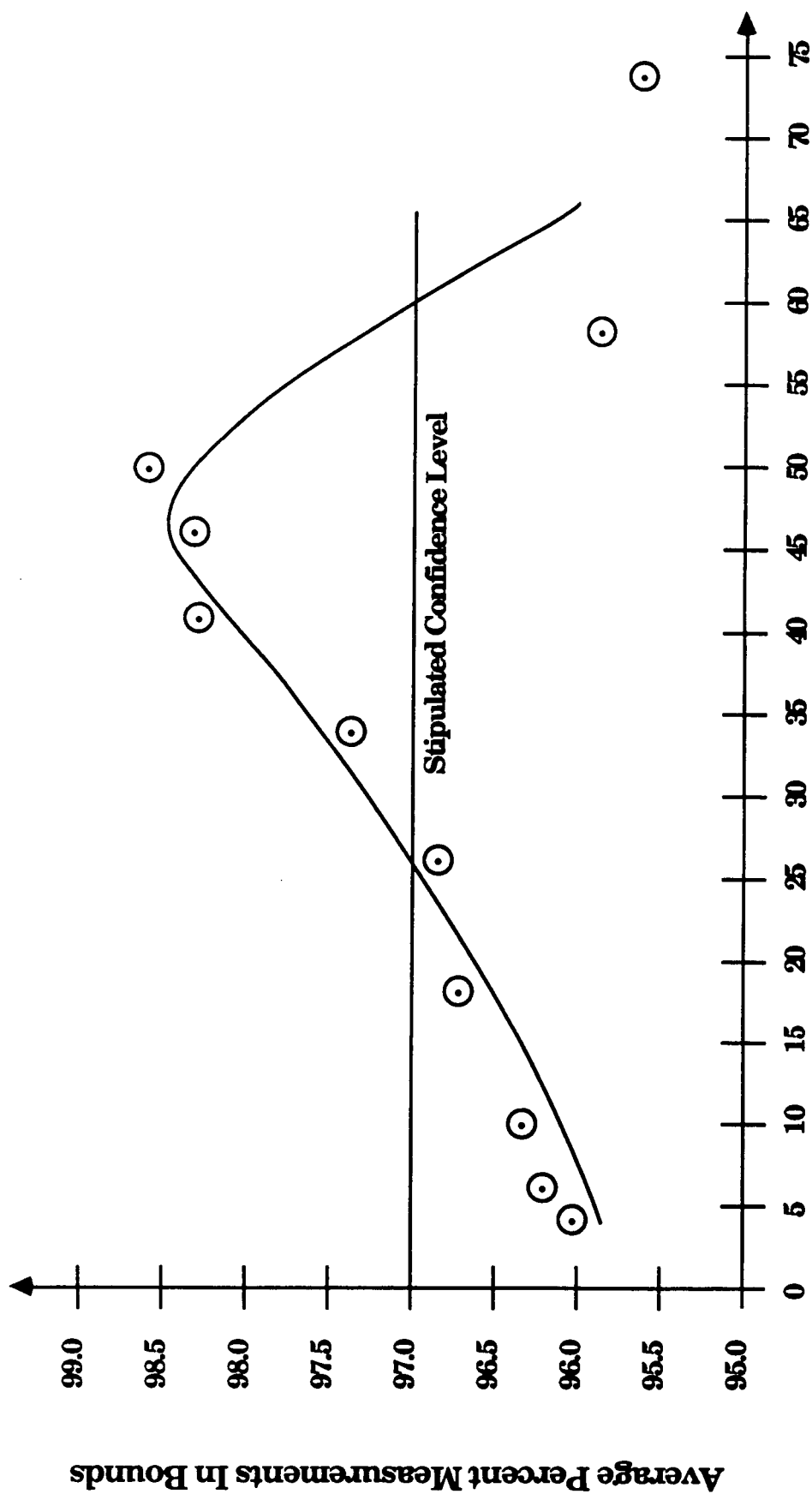
Figure 3.2-1 Experienced Confidence Level Sensitivity to Processing Parameters

-0.16 percent, relative to the stipulated or baseline confidence level.

It should be noted, however, that close scrutiny of the graphical measurements of Figures 3.2-2 through 3.2-12 reveal that mission 61-C has one measurement with far less data resolution than the others. Removal of this questionable measurement only affects the lower confidence bound profiles in a narrow band around the time periods t=-18.70 and t=-17.07 seconds and substantially leaves the average percent of in-bound measurements unaltered. This provides further evidence of the unbiased mean estimator efficiency.

Similar spectral results were obtained from the ET LH_2 Ullage Pressure and SSME LPFTP Discharge Temperature measurements. By comparison, the RSS Safe-and-Arm Device is of particular interest from a proof-of-concept standpoint in that it represents a binary on-off signal. Statistically, this type of measurement can only produce variation during its switch-over cycle whenever the phase shift timing is different between the various measurements. This aspect was not analyzed in this Phase I effort because it would have required the specification of infinitesimal Δt distances for the image point processing parameters, as well as the parameter used to define the adjacent distance between t_j confidence bound profile evaluations. Nevertheless, it is a surprising accomplishment to statistically provide such confidence of measurement samples utilized in this proof-of-concept effort.

Figure 3.2-2   2 Right And Left Side Image Points

Figure 3.2-3   3 Right And Left Side Image Points

Figure 3.2-4   5 Right And Left Side Image Points

28

Figure 3.2-5   9 Right And Left Side Image Points

29

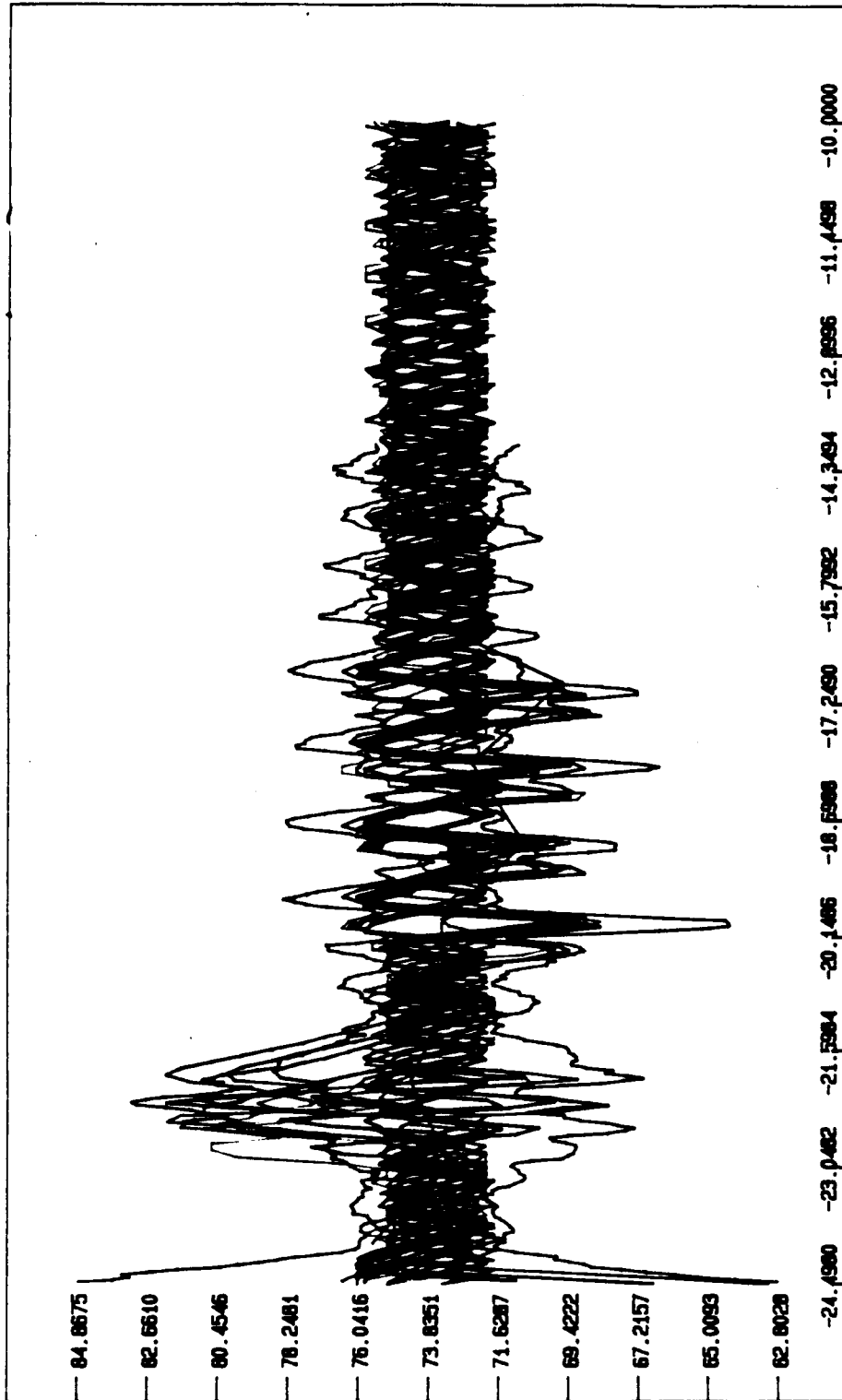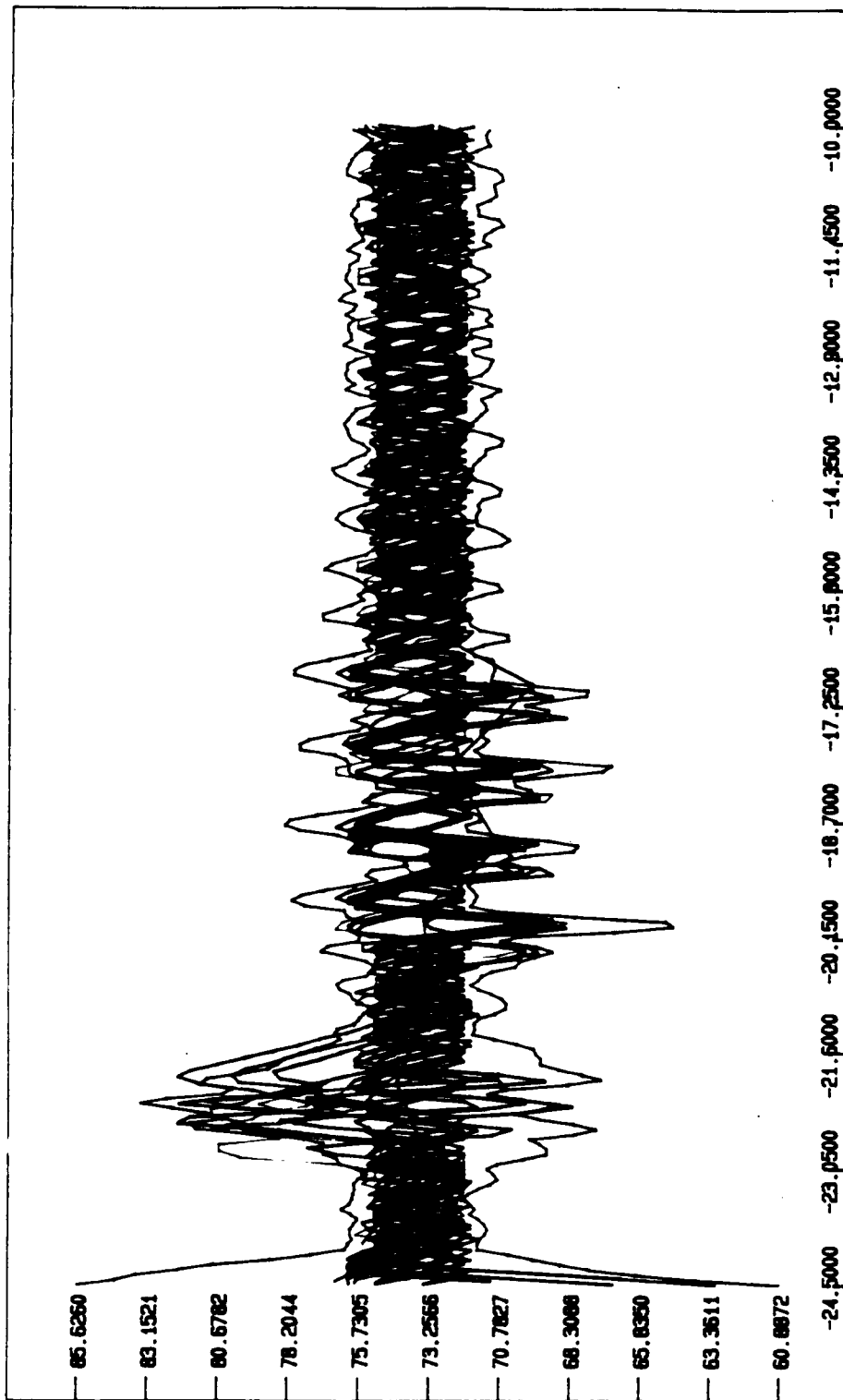Figure 3.2-6   13 Right And Left Side Image Points

Figure 3.2-7   17 Right And Left Side Image Points

31
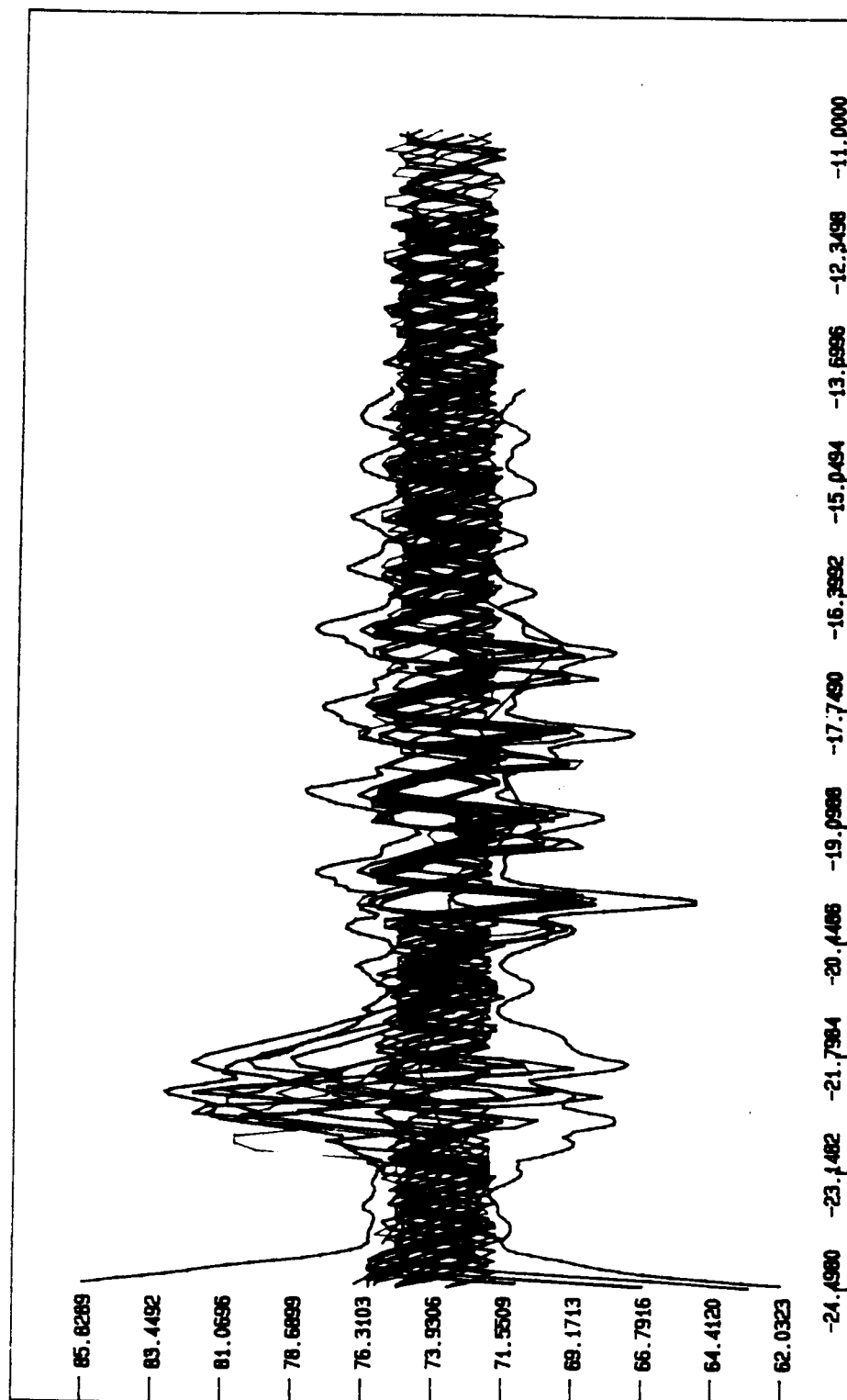
Figure 3.2-8   21 Right And Left Side Image Points
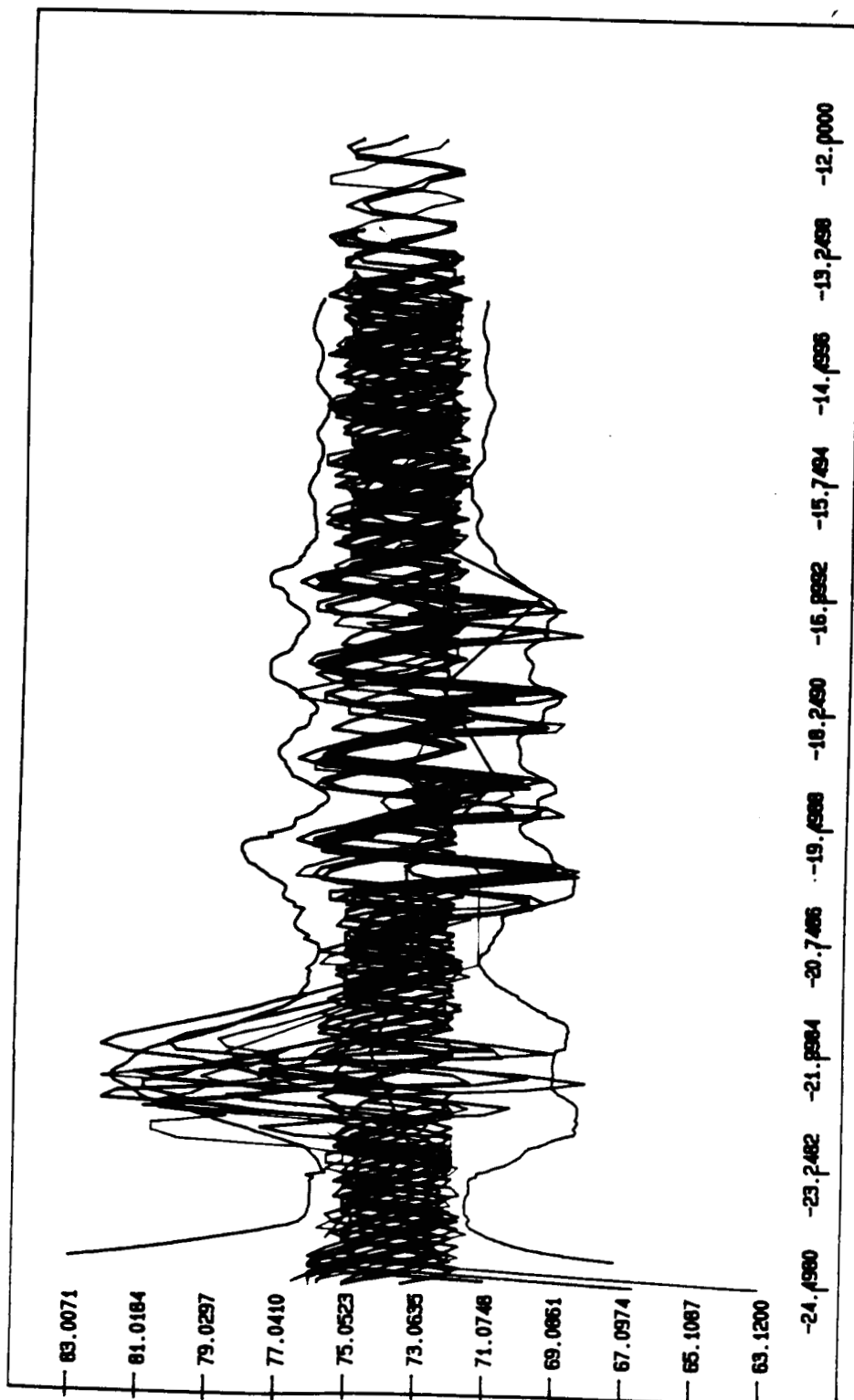
Figure 3.2-9   23 Right And Left Side Image Points
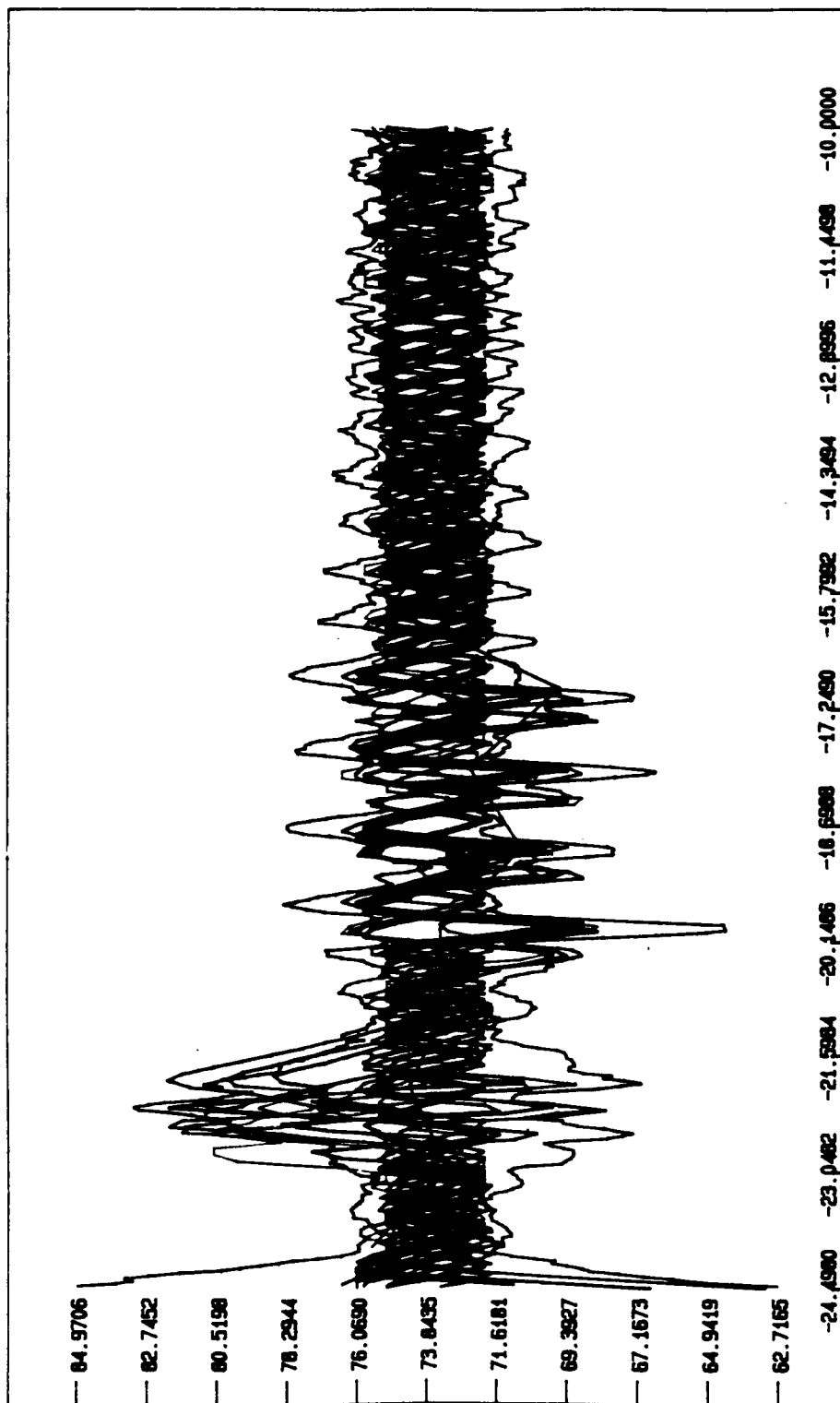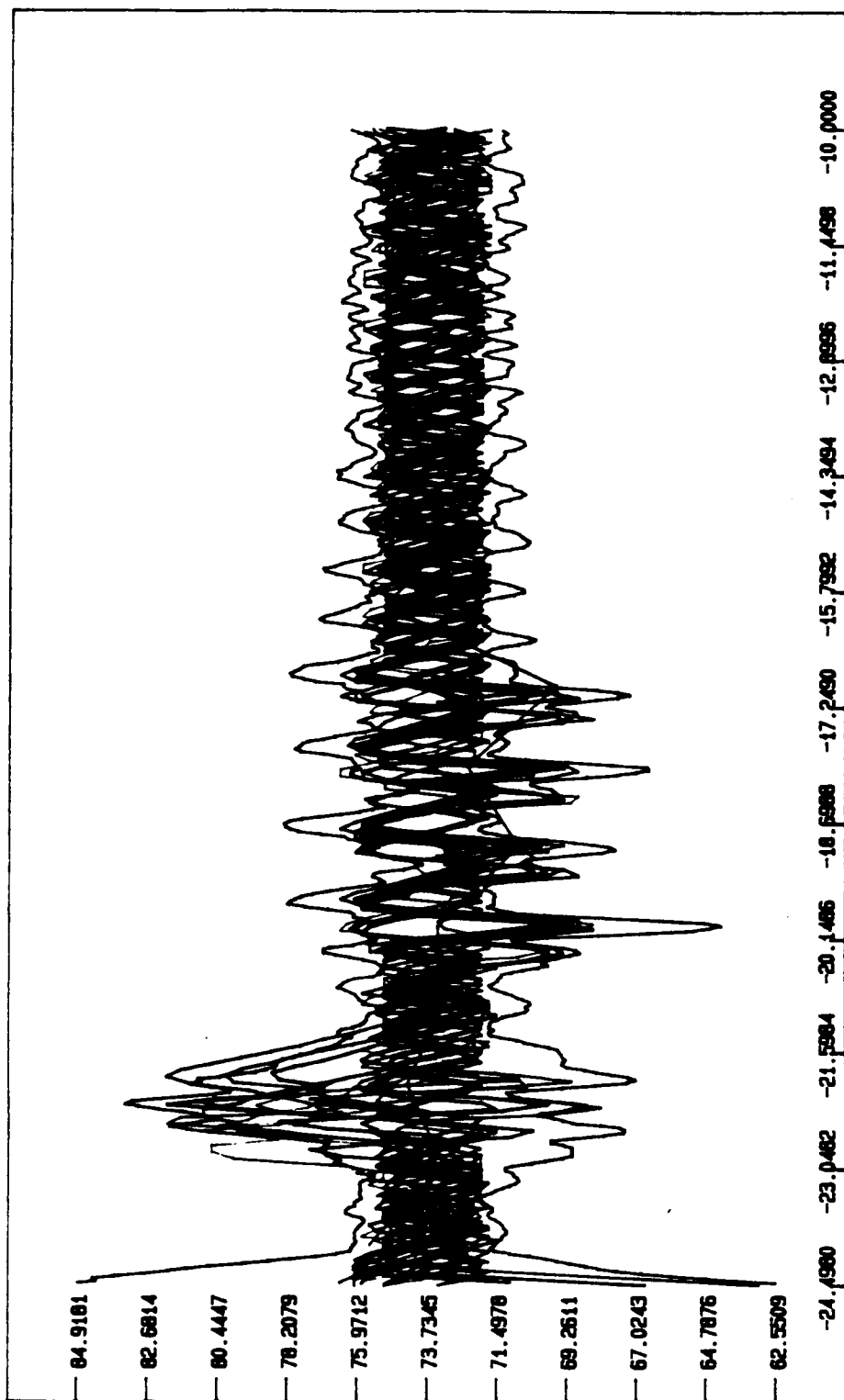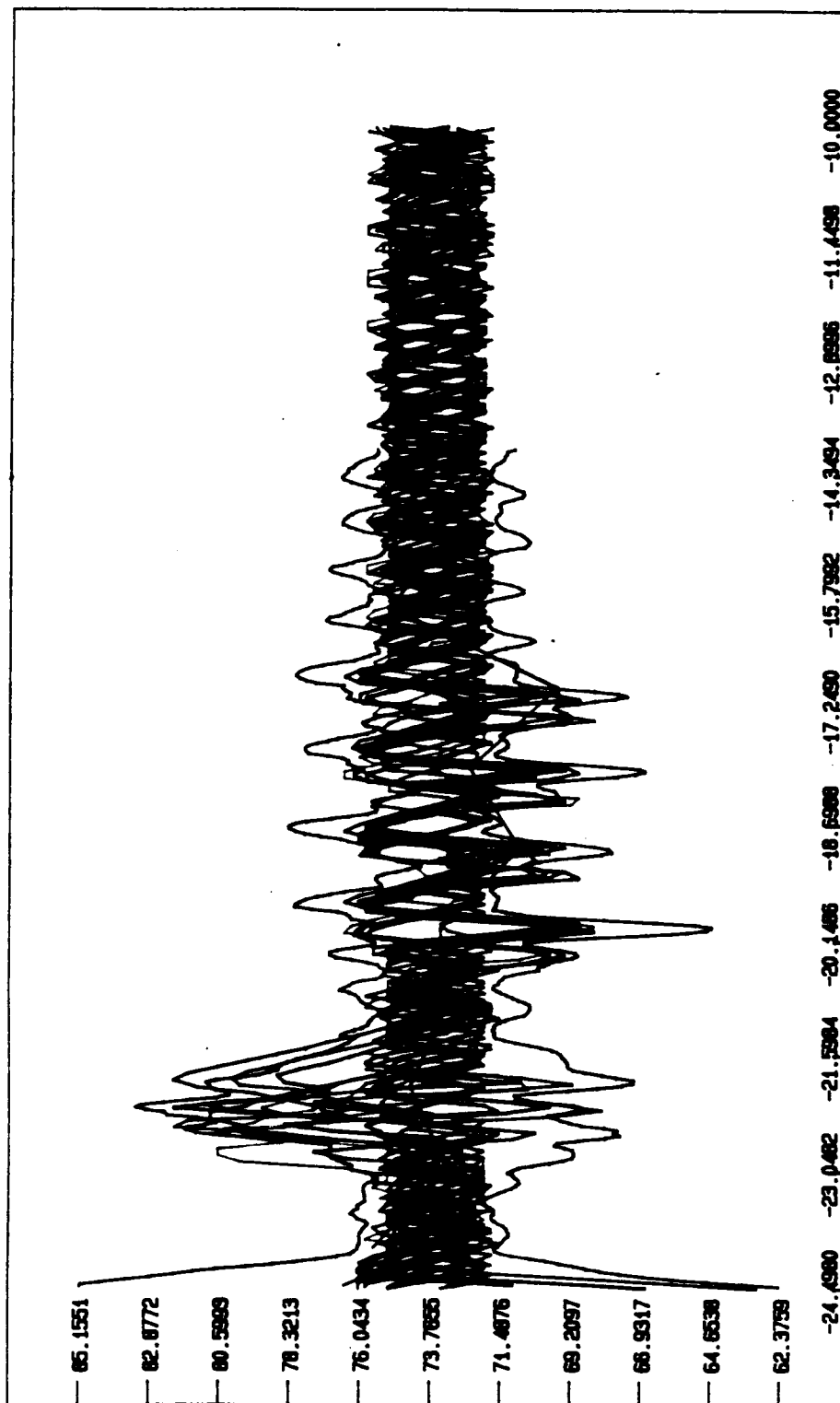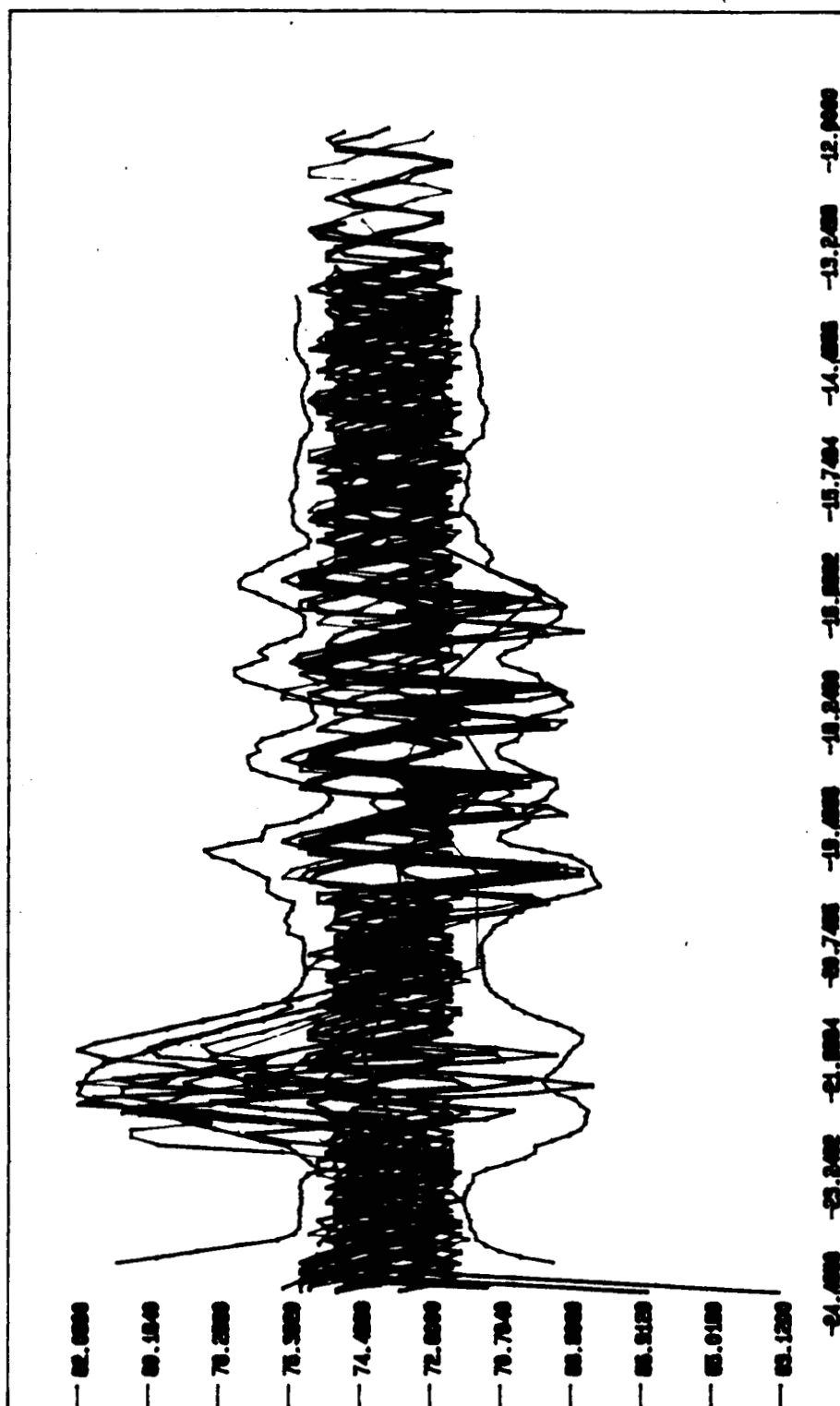
33
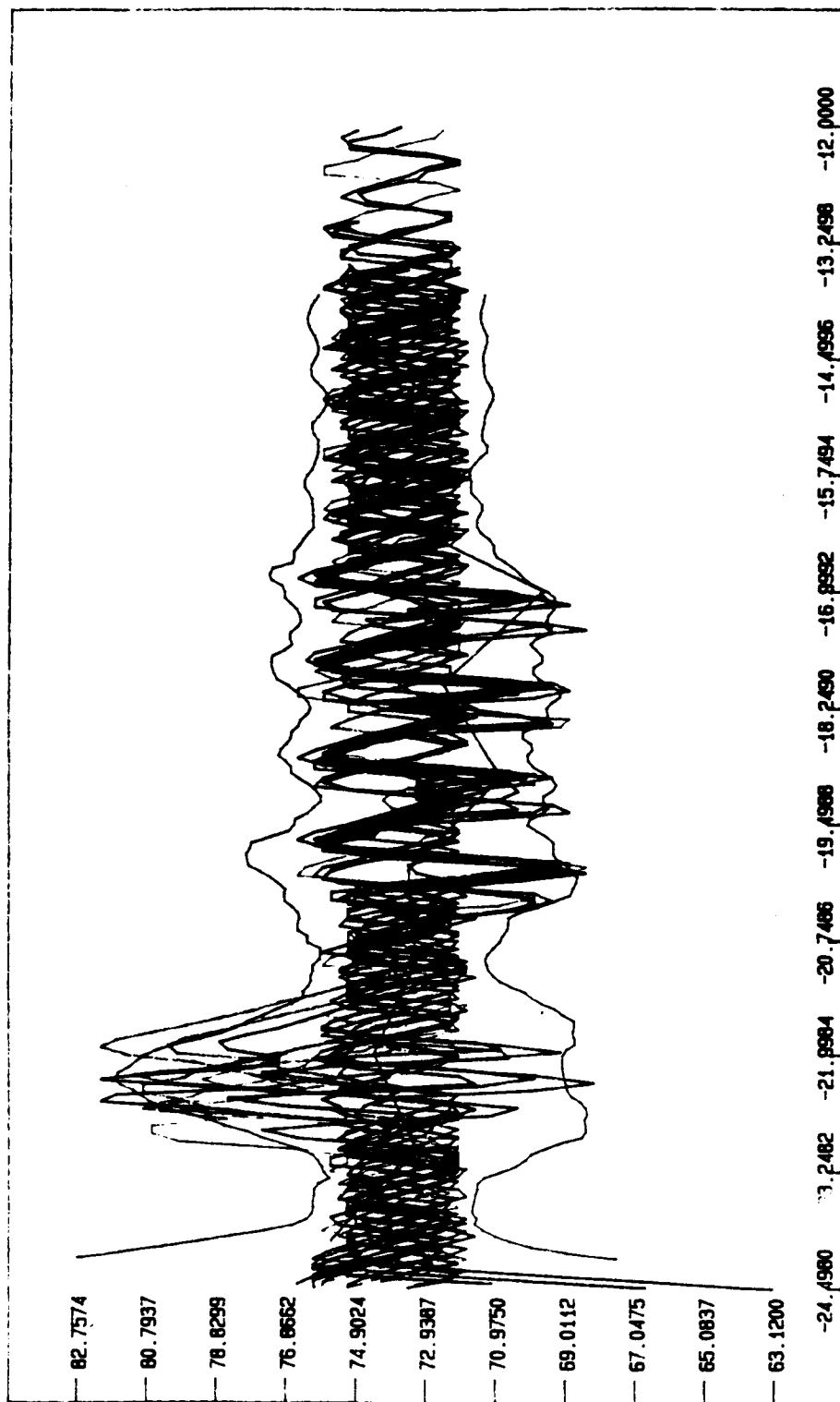
Figure 3.2-10   25 Right And Left Side Image Points

34
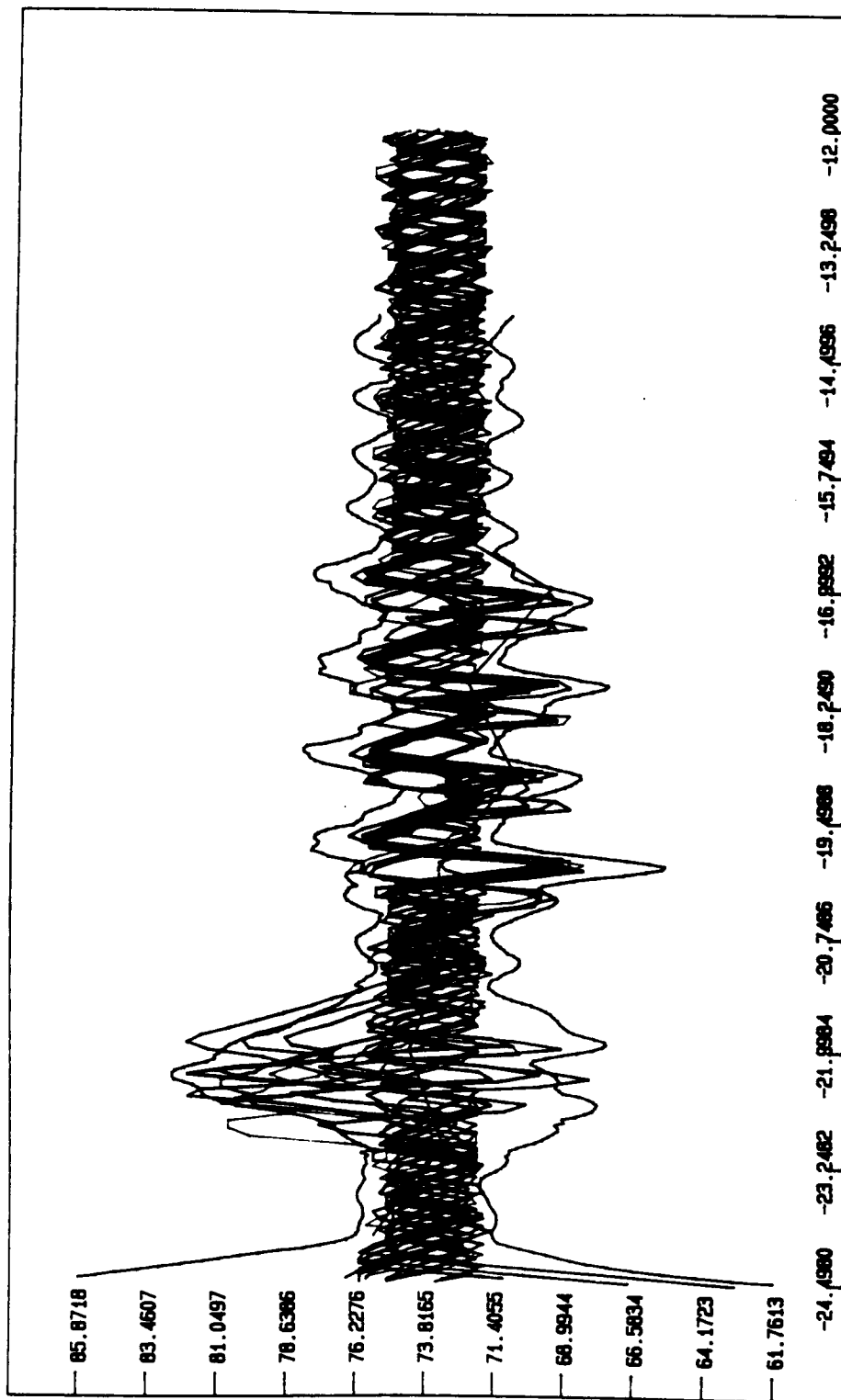
Figure 3.2-11   29 Right And Left Side Image Points

35
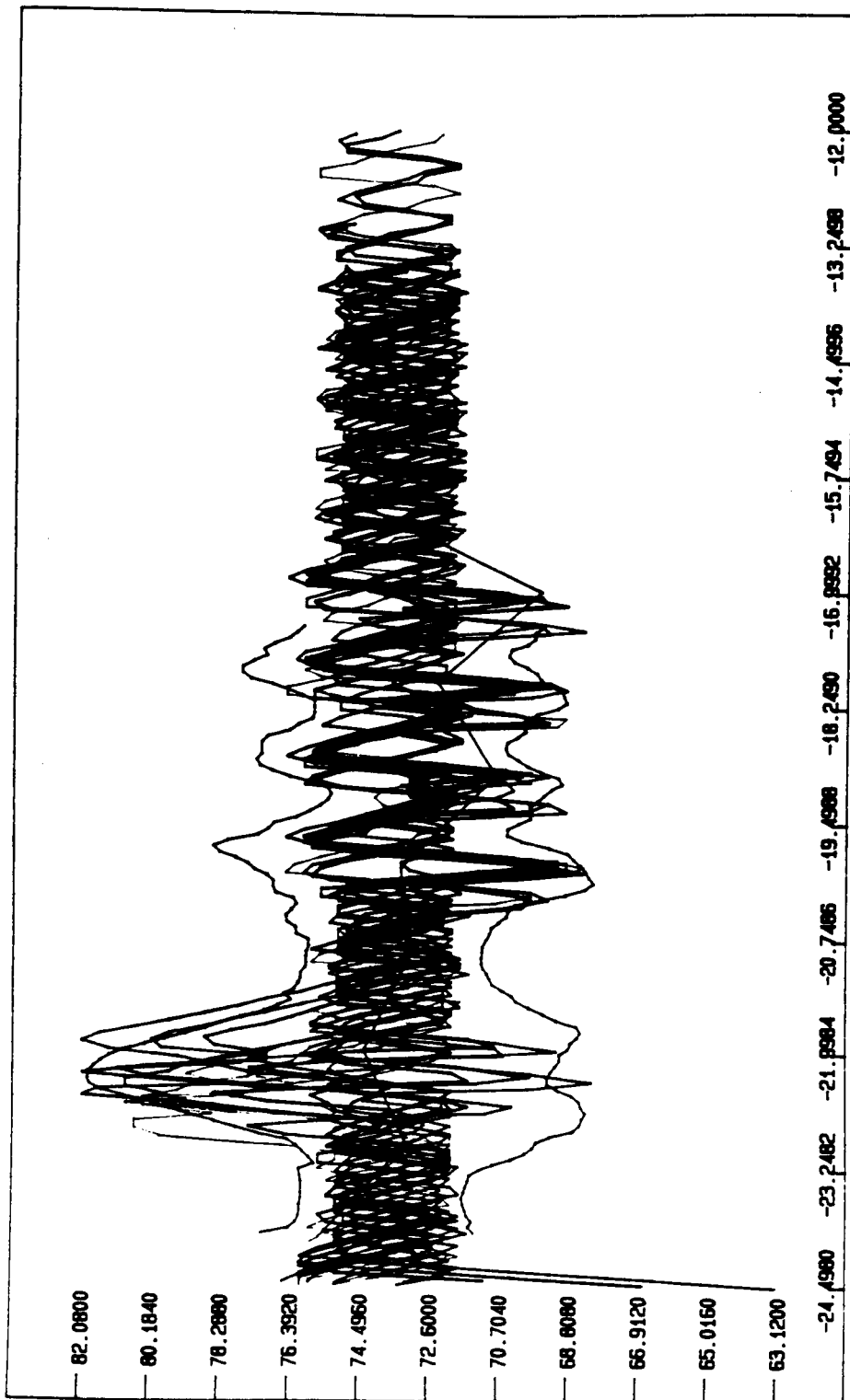
Figure 3.2-12   37 Right And Left Side Image Points

3.3 PROTOTYPE SOFTWARE COMPOSITION

Having demonstrated the statistical adequacy of experienced confidence level predictions and profile shaping parameter sensitivity, this section addresses the prototype software composition.

As with most prototype software development, there existed two overlapping and conflicting concerns: the desire to maximize code efficiency and the necessity to achieve recognized software clarity through relational code structures. To overcome these primitive code development concerns, the concept of establishing a subjective code quality metric was introduced. In this context, the term metric is defined as a measure of the extent or degree to which a given program segment (subroutine) exhibits each of the above concerns. This subjective code development framework also aided the mathematical formulation task in that algorithm formulation refinements could be correlated to the implemented software efficiency and clarity.

At this point, attention is directed toward the preprocessing and real-time software components introduced in the concept overview of Section 2.1. As indicated in this section, the preprocessing software component is responsible for data reduction, segmentation and data conditioning tasks needed to reduce the LCC measurements of a given type into a usable data format for statistical data assessment. It is this data conditioning mechanism that gives the overall process its efficiency in such a demanding computational environment.

Figure 3.3-1 provides a summary of the module level interaction used in this prototype software implementation to structure the above noted processing requirements.

DATA REDUCTION

GIVEN:
Raw LCC Measurement
Coordinate Data of a
Specified Measurement Type

Check File
(CKFILE)

- Verifies Data Readability
- Selects Data within a Specified Time-Interval
- Reconstructs Full-Rate Data

Parameter File Generation
(PARMGEN)

- Provides for Segment Processing Parameter Specification

Segment Identification
(SEGID)

- Establishes Individual LCC Measurement Segmentation

Composite Segment Identification
(COMPSEG)

- Constructs Composite LCC Data Segmentation from Individually Identified Segments Generated by SEGID

Update Segment Identification
(USPEGID)

- Reflects Composite Segment Characteristic on Individual Segments

Create Data Description
(CRDD)

- Defines Segment Time Domain Data Attributes

Ⓐ

Figure 3.3-1 Overview of Software Module Implementation

38

DATA CONDITIONING

- Selects Candidate Segment Data from CRDD

- Computes Governing Periodicity Equation Coefficients

- Identifies Optimum Constituent Frequency Root Search Pattern

- Evaluates Minimum Value of Periodicity Equation in a Closed Interval

- Evaluates Periodicity Equation for Specified Frequency Values

- Utilizes a Combination of Radial Pendulum Iteration and Wegstein's Accelerating Convergence Algorithm to Compute Closest Frequency (Root) in a Specific Direction

- Determines Frequency Whenever Two Values of the Periodicity Equation have Different Signs

- Prepares the Data for the ORTHF Routine

- Computes Scaling Coefficients (Note: Same Routine as Above)

(B)

(A)

Constituent Periodicity Determination (CONPD)

Statistical Orthonormalization Smoothing (Propriety Algorithm) (ORTHF)

Governing Periodicity Equation Root Search (ROOTS)

Minimum Functional Value (MINFV)

Governing Periodicity Equation Deviation (GRED)

Wegstein Accelerated Convergence (WEGAC)

Zero Root Evaluation (ZERO)

Scaling Coefficient Preprocessing (SCLPRP)

Statistical Orthonormalization Smoothing (Propriety Algrithm) (ORTHF)

Figure 3.3-1   Overview of Software Module Implementation (Cont'd)

DATA STATISTICAL ASSESSMENT

Ⓑ

Aggregated Segment Evaluation (ASEGS)

- Determines Unbiased Mean Estimator needed to Establish GCS "Distribution-Free" pdf

Aggregated Segment Bound (ASBND)

- Forms GCS Distribution and Evaluates Upper and Lower Confidence Bound Profiles for Specified Confidence Level

Historical Mission Data Base

Contains Segmented Data:
- Frequencies, W_j's
- Scaling Coefficients, A_j's, B_j's
- Segment Type I.D.
- Beginning and End Point Attributes

Trend Report Data Base

REAL-TIME MISSION DATA TRENDING

Real-Time Trend Assessment (REALTIME)

- Compares New Trend Data with Confidence Bound Profiles and Identifies Time-Varying Trend Anomalies
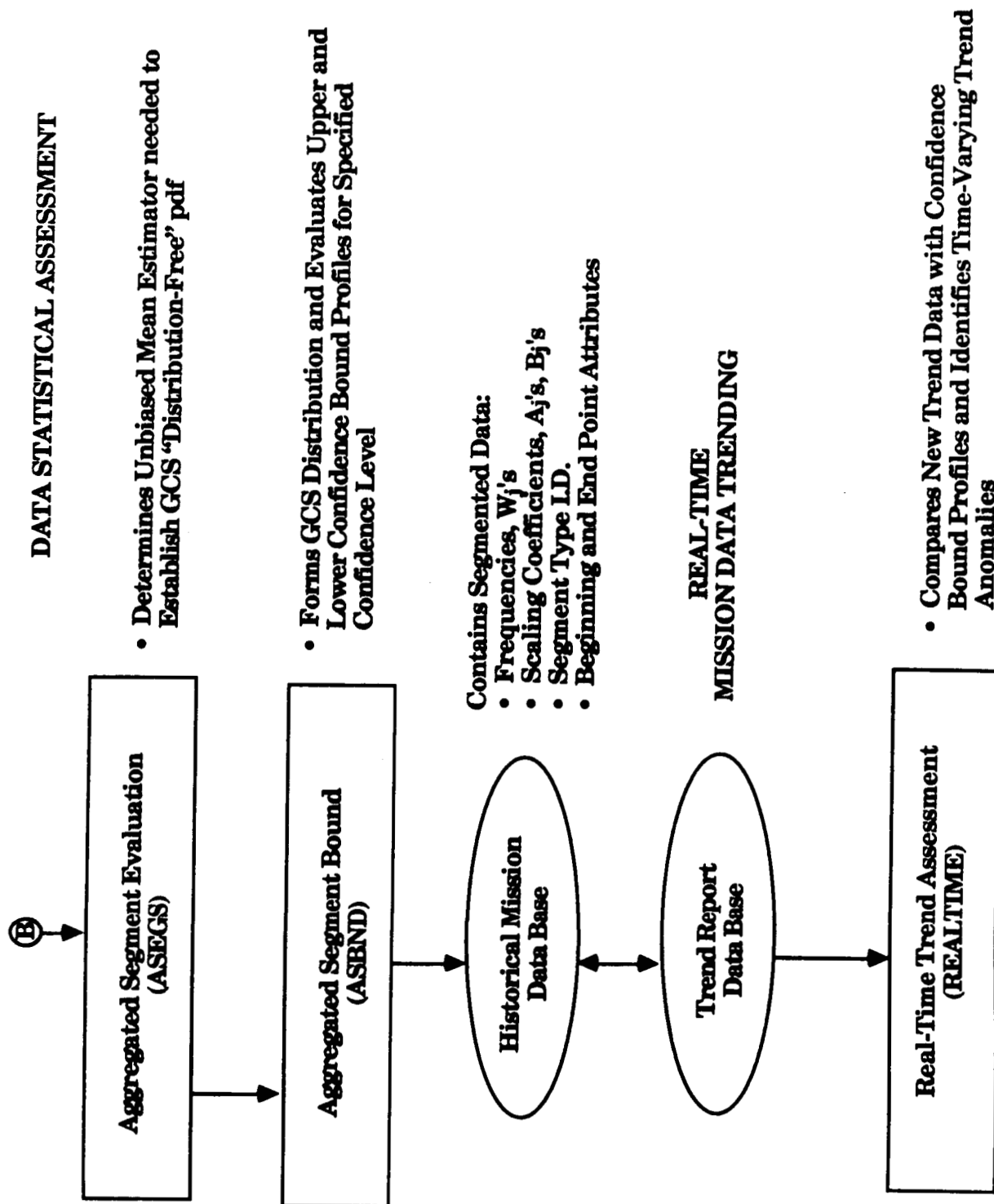
Figure 3.3-1   Overview of Software Module Implementation (Cont'd)

3.4 PROOF-OF-CONCEPT TRENDING RESULTS

During this phase of the effort, applicable confidence level segment profiles were generated for the four selected LCC measurement types from the five missions prior to 51-L. Then attention was directed towards determining the confidence level sensitivity of each segment profile. To assess this sensitivity, the average percent of measurements within the respective confidence bound profiles was determined. Graphical substantiation of this sensitivity was also provided by looking at the confidence interval distance as a function of segment time and confidence level uncertainty for each applicable data segment and measurement type.

Once this activity was completed, the relevant LCC trending reports (graphic displays) associated with the applicable STS-26 measurements were induced and analyzed for trend anomalies. Indicative trending results for selected STS-26 SRB APU Turbine Speed, ET LH_2 Ullage Pressure and SSME LPFTP Discharge Temperature measurements are graphically presented in Figures 3.4-1 through 3.4-3 for 95 percent confidence level bound profiles. The image point parameter specification used for each of these figures is provided below:

1) Delta distance from t_j equals 0.01562.
2) Distance between adjacent t_j's equals 0.03125.
3) Number of assimilated image points per measurement equals ten.

In Figure 3.4-1, the peculiar (alternating) phase shifts of the selected STS-26 measurement relative to the upper and lower confidence bound profiles appears to reflect a shifted turbine speed start-up sequence on the STS-26 mission. Another related phase shift phenomenology is illustrated in Figure 3.4-2. In this instance, the upper and lower ullage pressure confidence profiles appear to have a shorter period than the selected STS-26 measurement. This spatiotemporal phenomenon appears to be related to tank time/temperature differences between individual mission launch environments. Another point to be considered regarding the ullage pressure assessment is that only 15 measurements were available, i.e., three measurements per mission, over five missions.

In reference to the SSME LPFTP Discharge Temperature trending assessment of Figure 3.4-3, it should be noted that both Channel A and B data (i.e., with markedly different data resolutions) were used in this STS-26 Channel B characterization. By comparison, the pursuant RSS Safe-and-Arm Device trending assessment of Figure 3.4-4 shows no upper and lower confidence bound profiles because the binary on-off signal (measurements) lack statistical variance. See Section 3.2 comments for added clarification.

41
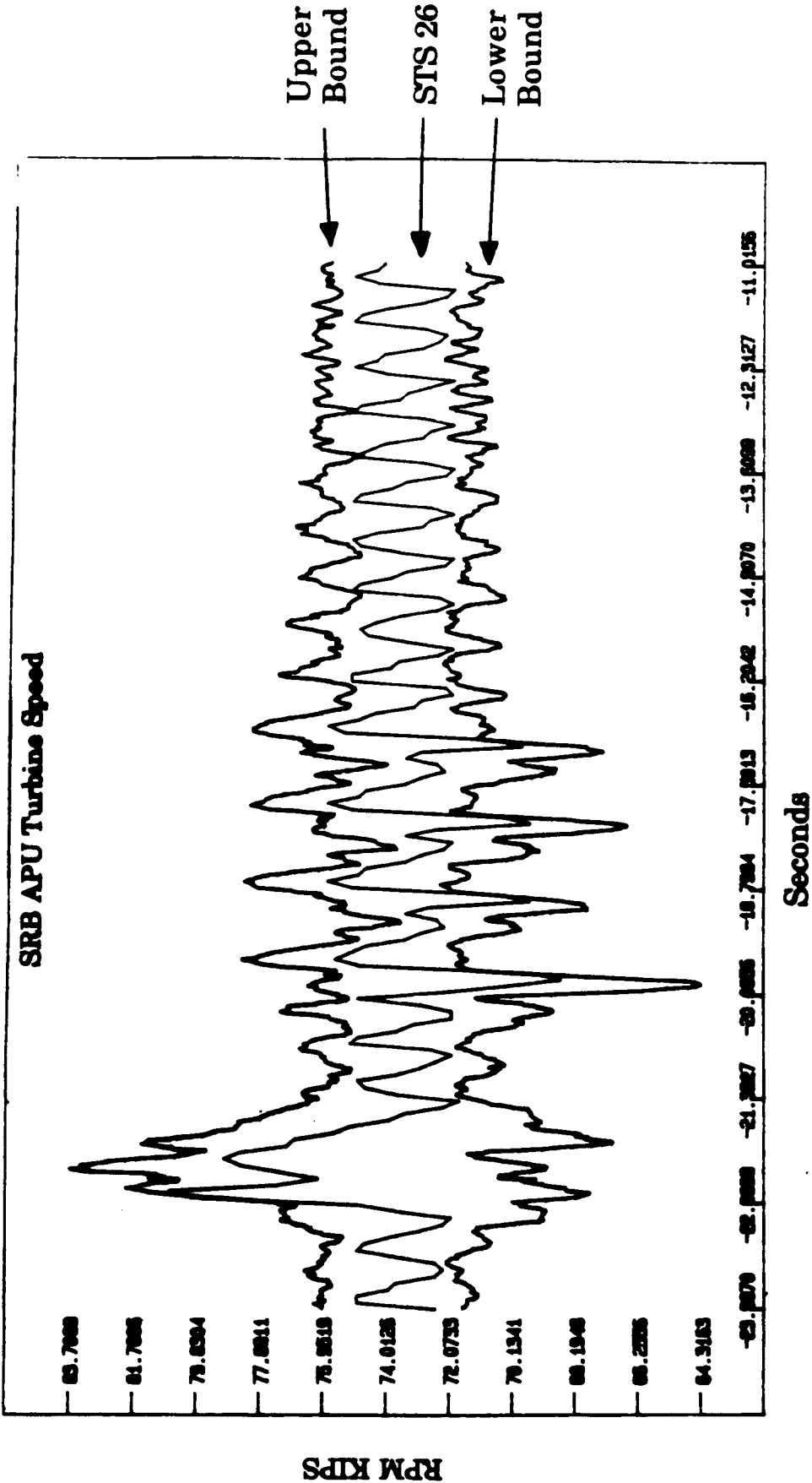
STS-26 Trended Data

LCC Measurement: SRB APU Turbine Speed
Confidence Level: 95%

Figure 3.4-1   STS-26 Trended Data for SRB APU Turbine Speed

STS-26 Trended Data

ET LH$_2$ Ullage Pressure

ORIGINAL PAGE IS
OF POOR QUALITY

Upper Bound

STS 26

Lower Bound

PSIA

Seconds
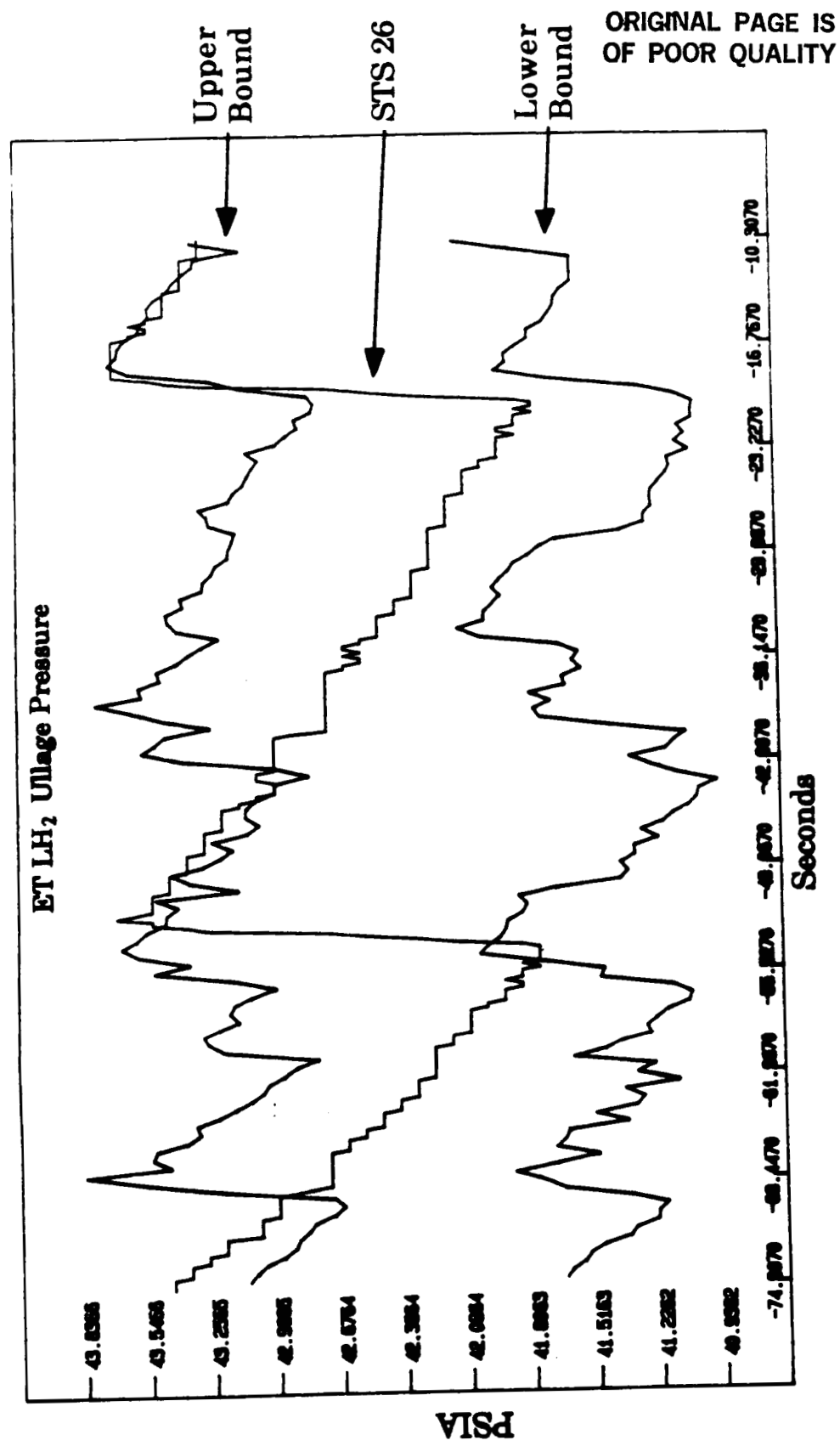
LCC Measurement: ET LH$_2$ Ullage Pressure
Confidence Level: 95%

Figure 3.4-2   STS-26 Trended Data for ET LH$_2$ Ullage Pressure
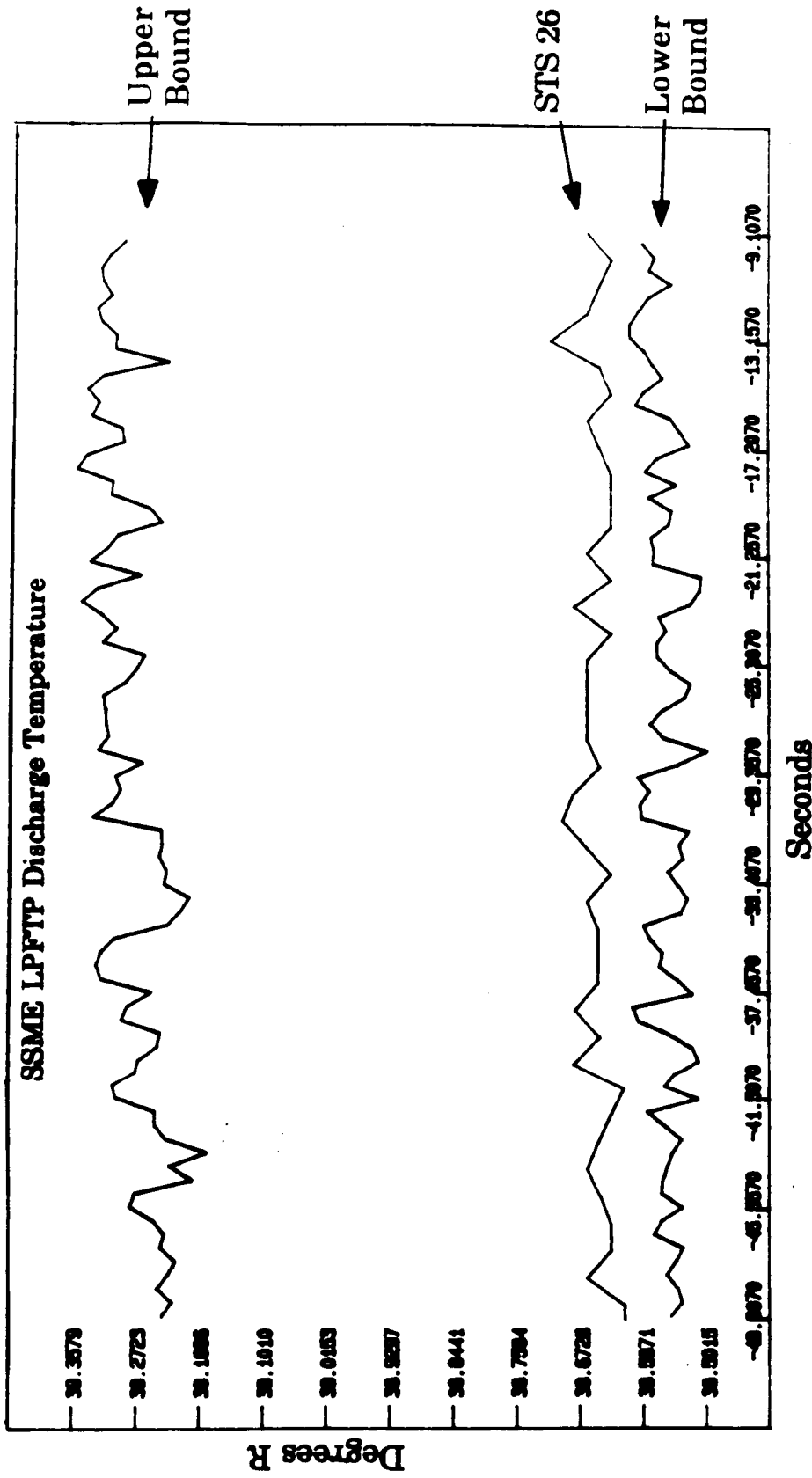
STS-26 Trended Data

LCC Measurement: SSME LPFTP Discharge Temperature
Confidence Level: 95%

Figure 3.4-3   STS-26 Trended Data for SSME LPFTP
Discharge Temperature

44

Generated Statistical Confidence Bounds

Upper & Lower
Bounds As Well As
Arm Measurements

Results From Lack
Of Measurement
Variance

RSS Safe and Arm Device

0-1 Signal

Seconds

Redlines:
Min   Max
ON   NA
T-4 min 55 sec to T-10 sec

LCC Measurement: RSS Safe and Arm Device
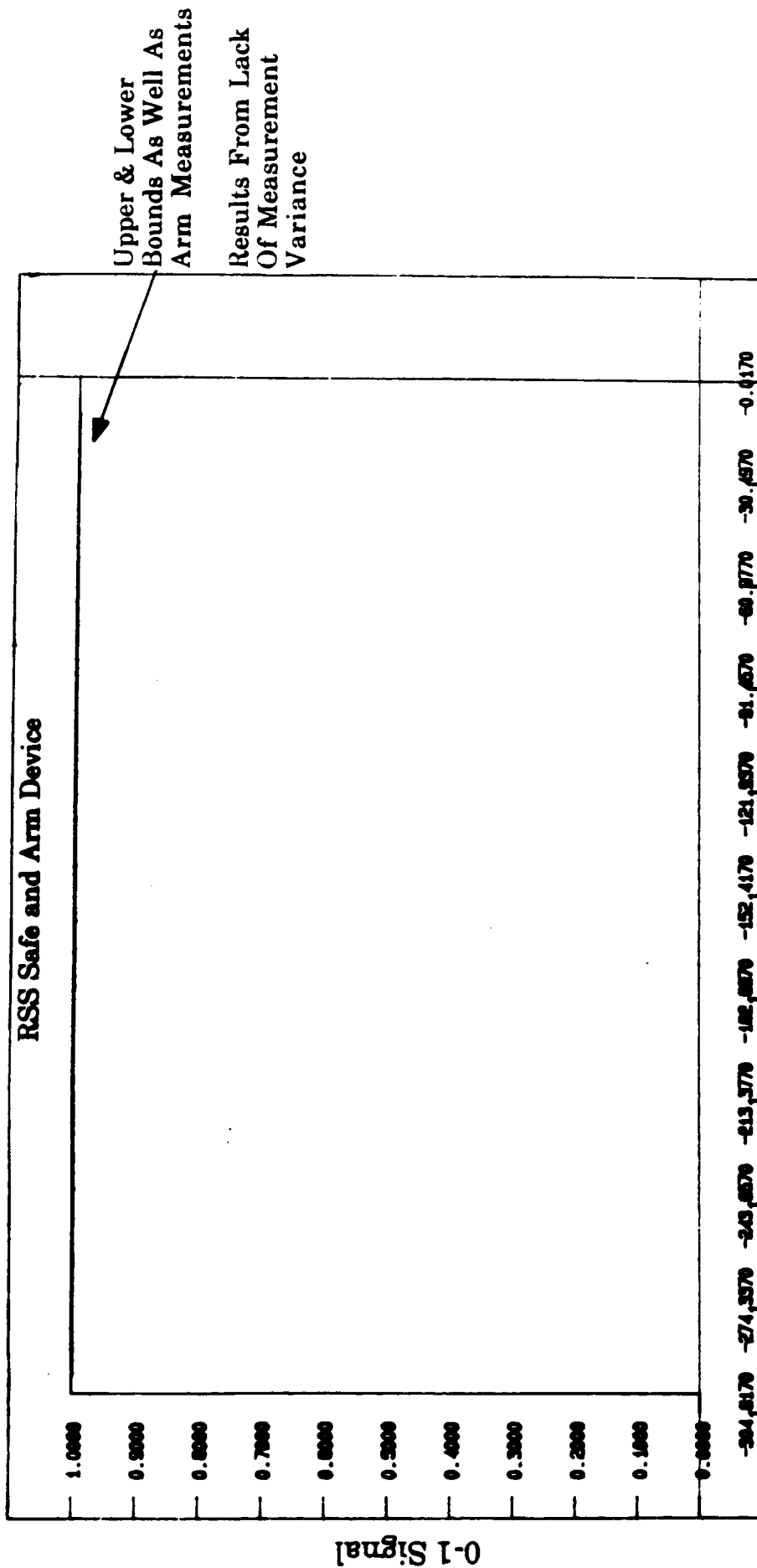Confidence Level: 95%

Figure 3.4-4   Generated Statistical Confidence Bounds for
RSS Safe and Arm Device

45

Further illustrative results are provided in Figures 3.4-5 through 3.4-7 which show the upper and lower confidence bounds superimposed on the generating measurements. It is of particular interest to note the in-phase measurement grouping of Figure 3.4-5 within the approximate time interval -19.8 to -15.5 seconds. Another interesting peculiarity is noted in Figure 3.4-7 where the Channel B measurement of Main Engine One associated with mission 61-A appears as a statistical outlier.

To check the confidence level sensitivity associated with the measurements depicted in Figures 3.4-5 through 3.4-7, equidistant time cuts with 0.250 spacing were constructed over the indicated confidence profile intervals. Then these intervals were evaluated and the averaged proportion of in/out measurements were used as the actual confidence level. These values were 96.3, 95.8, and 95.01, respectively, for the measurements of Figures 3.4-5 through 3.4-7.

4.0 CONCLUSIONS AND RECOMMENDATIONS

4.1 CONCLUSIONS

It has been demonstrated that the GCS is a "distribution-free" pdf with explicit asymptotic conformity to the central limit theorem. In addition, the capability to generate highly accurate upper and lower confidence bound profiles from this methodolody has been established. Consequently, it can safely be stated that this proof-of-concept is substantiated.

4.2 RECOMMENDATIONS

Given that the concepts of an LCC measurement population and the selection of random measurement samples are basic to this inductive LCCTA methodology, it is recommended that the Phase I effort be proceeded by adding expanded LCC measurements of selected types from previous missions. To help reveal any underlying idiosyncrasies embedded in these measurements due to peculiar operational phenomenologies or design changes, it is further recommended that the aggregated bound variations from mission to mission be assessed to identify any mission/measurement drivers.

In addition, it is recommended that the Phase I prototype software be refined and integrated into a self-contained package to support the above processing requirements. This activity should also include the preparation of full-user software and methodology documentation to permit potential integration of this trending technique into other systems and applications.

Generated Statistical Confidence Bounds

	Min(RPM)	Max(RPM)
Redlines:		
T-26 to T-21 sec	10000	85000
T-21 to T-16 sec	66200	85000
T-16 to T-0 sec	66200	77800

LCC Measurement: SRB APU Turbine Speed
Confidence Level: 95%
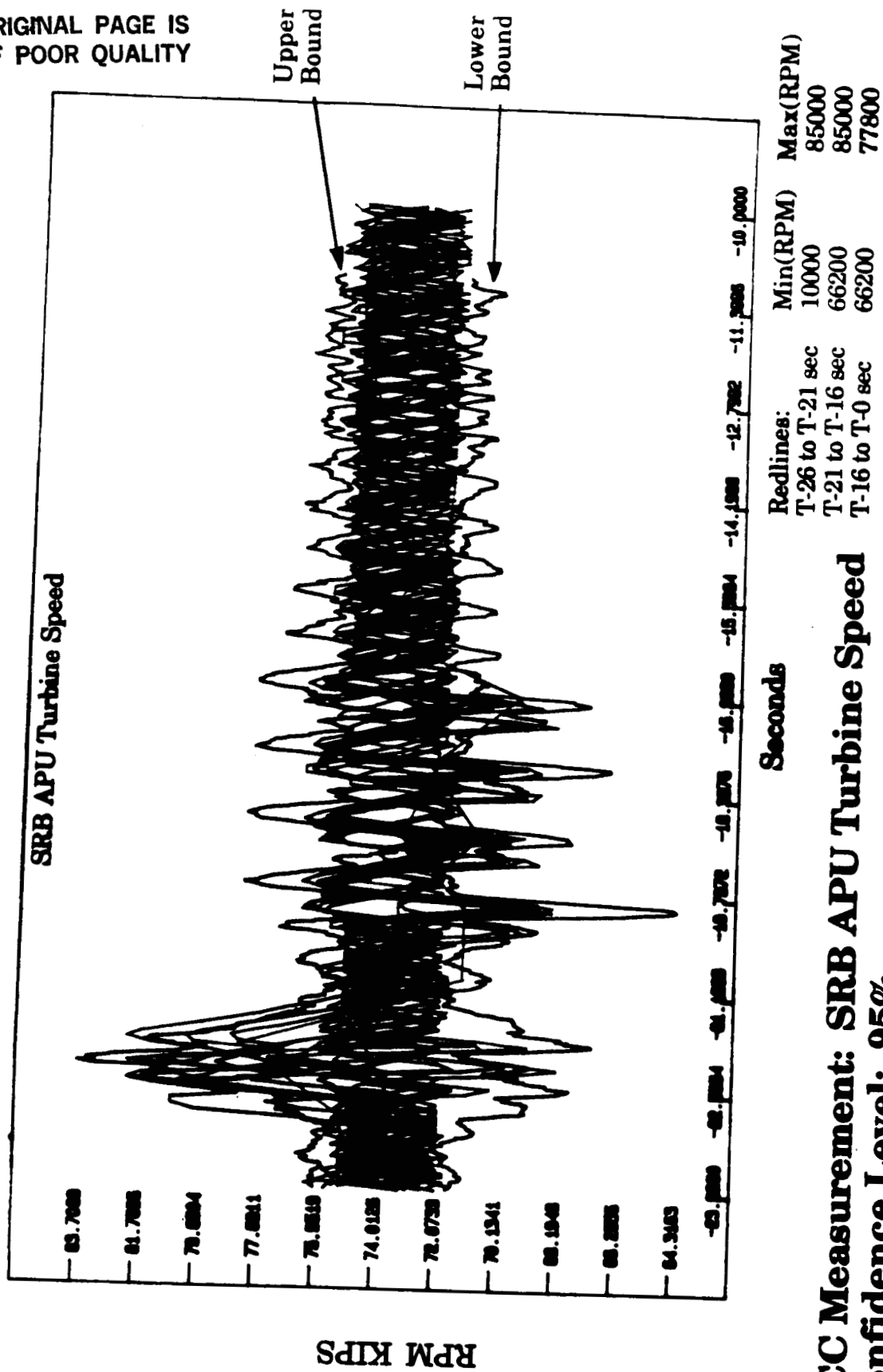
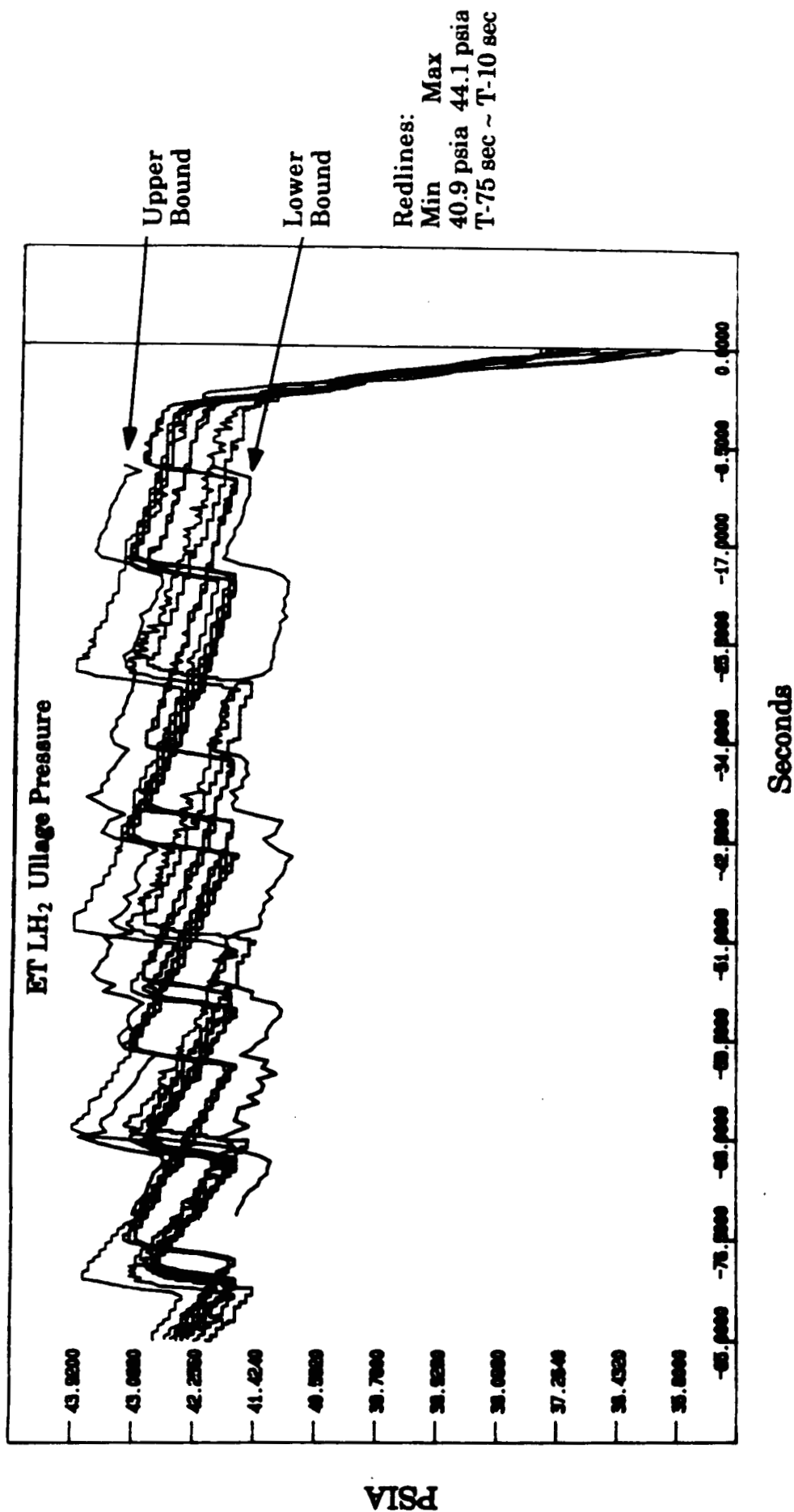Figure 3.4-5   Generated Statistical Confidence Bounds
SRB APU Turbine Speed

Generated Statistical Confidence Bounds

ET LH$_2$ Ullage Pressure

Upper Bound

Lower Bound

Redlines:

Min   Max
40.9 psia   44.1 psia
T-75 sec ~ T-10 sec

PSIA

Seconds

LCC Measurement: ET LH$_2$ Ullage Pressure
Confidence Level: 95%

Figure 3.4-6   Generated Statistical Confidence Bounds
ET LH$_2$ Ullage Pressure

48

Generated Statistical Confidence Bounds

LCC Measurement: SSME LPFTP Discharge Temperature
Confidence Level: 95%

Figure 3.4-5   Generated Statistical Confidence Bounds
SSME LPFTP Discharge Temperature

49
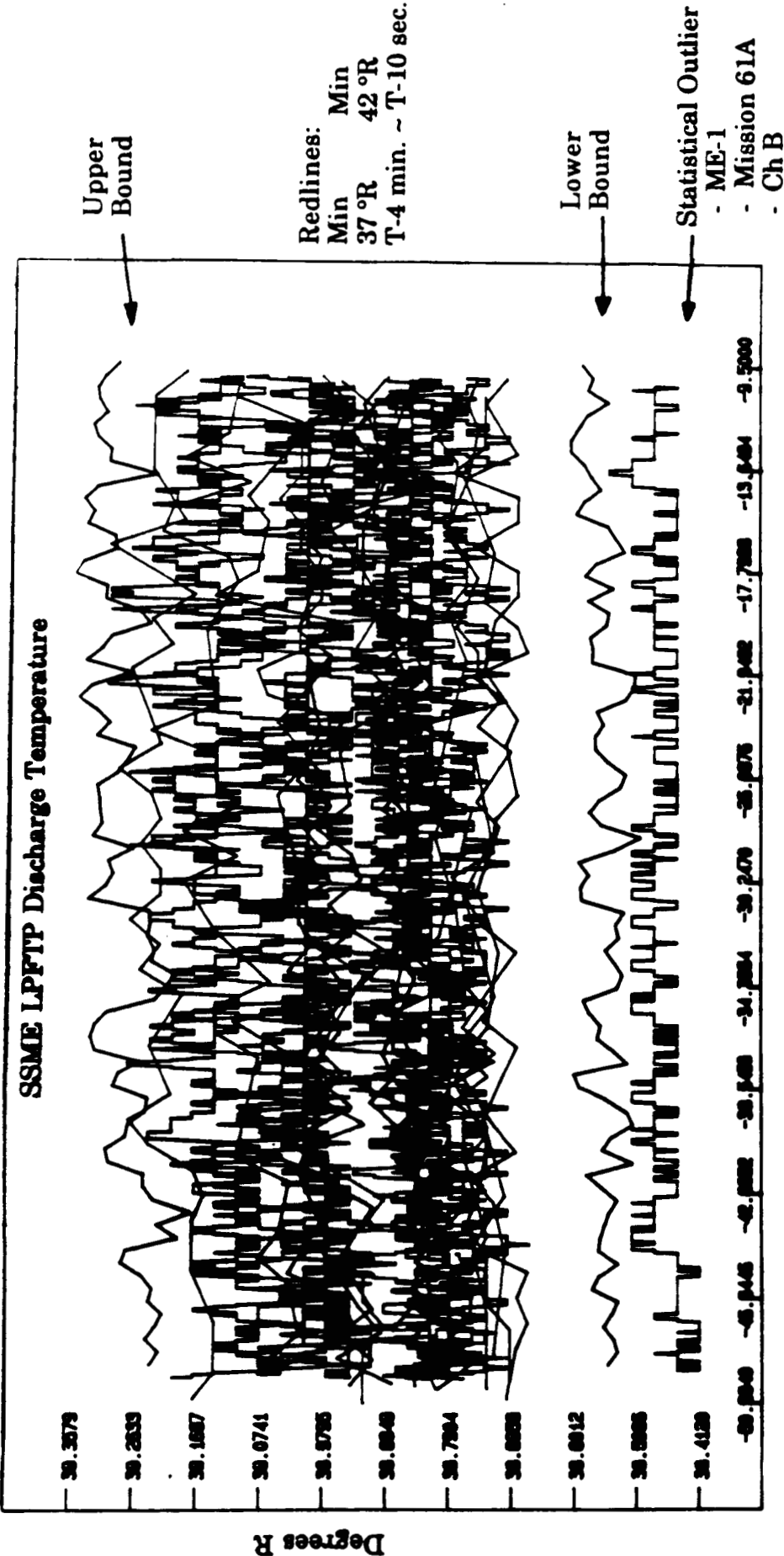
APPENDIX A

STATISTICAL FORMULATION

Appendix A - Statistical Formulation
Table of Contents

APPENDIX A

STATISTICAL FORMULATION

A.1 Historical Overview

The primary mathematical focus of the Phase I proof-of-concept effort was directed toward the development of a "distribution-free" characterization of the time-varying statistical confidence bounds associated with historical Launch Commit Criteria (LCC) measurement data profiles. Once developed, this capability provides the opportunity to assess the occurrence of potential measurement data anomaly trends and other abnormal operational conditions relating to erratic data amplitude and phase shift variations.

Since the statistical data variation contained in historical successful mission measurements must be related to aggregated time-varying probability density functions, the attainment of these distribution parameters are of paramount importance to the stability of the resulting confidence bound assessment. The capability to characterize these distribution parameters is set in the early statistical investigations conducted by Gram (see Reference 5). In this basic research, Gram showed that an arbitrary frequency distribution, defined on an infinite interval, can be generated in terms of the sum of independent frequency distributions. It was also expounded in this original work that the normal symmetrical Gaussian error curve is but a special case of a more general system of skewed frequency distributions which can be represented by an approximating series expansion.

It is also interesting from a historical perspective to note that Thiele confirmed this far-reaching idea in 1887. The substance of this confirmation was later published in a book entitled, "A General Theory of Observations" (see Reference 6). Among the added frequency distribution achievements of Thiele was the introduction into statistics of the notion of moments, which he gave the name "seminvariants." Subsequent statistical literature usage of this term is "semi-invariants." By means of these semi-invariants, Thiele illustrated the relationship between sample data histogram frequency approximations and the series expansion of Gram's arbitrary frequency distribution.

In addition to Thiele's work, the Swedish astonomer C. V. Charlier introduced an efficient computational methodology for expressing the various distribution parameters from Thiele's semi-invariants. This was accomplished by the utilization of both orthogonal functions and integral equations. Thus, this series expansion was designated as the Gram-Charlier Series (GCS) in the subsequent statistical literature of the time.

With this historical background, the specious question of why this series expansion has not received the attention it deserves in the current literature is more speculative. The most probable reason is the fragmentary and unsystematic manner in which it was developed.

A.2 Mathematical Background

To place this GCS analytical development in its proper orientation, a remarkable functional property between Hermite polynomials and the repeated differentiation of the Gaussian error curve,

$$\Psi(z) = \frac{1}{\sqrt{2\pi}} e^{\frac{-z^2}{2}}$$

exists. Denoting the various derivatives of $\Psi(z)$ by

$$\Psi^k(z) = \frac{d^k}{dz^k} (\Psi(z))$$

for $k=0, 1, 2, \ldots$, we obtain the following relations:

$$\Psi^{(0)}(z) = \Psi(z) = \frac{1}{\sqrt{2\pi}} e^{\frac{-z^2}{2}}$$

$$\Psi^{(1)}(z) = -z\Psi(z)$$

$$\Psi^{(2)}(z) = (z^2-1)\Psi(z)$$

$$\Psi^{(3)}(z) = -(z^3-3z)\Psi(z)$$

$$\Psi^{(4)}(z) = (z^4-6z^2+3)\Psi(z)$$

$$\vdots$$

Given that these derivatives are represented as the products of polynomials of z and the Gaussian error curve itself, it is only natural to inquire about the functional polynomial forms involved. As it turns out, these polynomials are known as Hermite's polynomials from the name of the French mathematician, Hermite, who first introduced the recursive formulation,

$$H_{n+1}(z) = zH_n(z) - nH_{n-1}(z)$$

for this class of orthogonal polynomials. Upon utilizing this recursive formulation with starting values,

$$H_0(z) = 1 \quad \text{and} \quad H_1(z) = z,$$

the successive polynomial forms:

$$H_2(z) = z^2 - 1$$

$$H_3(z) = z^3 - 3z$$

$$H_4(z) = z^4 - 6z^2 + 3$$

$$\vdots$$

are exhibited that are contained in the various derivatives of the Gaussian error curve.

From a numerical computation point-of-view, it is convenient to recursively use the expression

$$G_{n+1}(z) = \frac{1}{n+1}[zG_n(z) - G_{n-1}(z)]$$

together with the starting values

$$G_0(z) = \Psi(z) \quad \text{and} \quad G_1(z) = zG_0(z)$$

to obtain the general form

$$\Psi^{(0)}(z) = \Psi(z)H_0(z) = G_0(z)$$

$$\Psi^{(1)}(z) = -\Psi(z)H_1(z) = -G_1(z)$$

$$\Psi^{(2)}(z) = \Psi(z)H_2(z) = 2G_2(z)$$

$$\vdots \qquad \vdots \qquad \vdots$$

$$\Psi^n(z) = (-1)^n \Psi(z)H_n(z) = (-1)^n n! G_n(z)$$

of the Gaussian error curve derivatives.

Before addressing the generation of the GCS, it is necessary to discuss some important integral equation relationships between the Hermite polynomials and the various derivatives of $\Psi(z)$ or $\Psi^{(0)}(z)$. To this end, consider the following series of functions:

$$\Psi^{(0)}(z), \ \Psi^{(1)}(z), \ \Psi^{(2)}(z), \ \Psi^{(3)}(z), \ \ldots$$

$$H_0(z), \ H_1(z), \ H_2(z), \ H_3(z), \ \ldots$$

where

$$\Psi^{(n)}(z) = (-1)^n \Psi(z) H_n(z)$$

with

$$\lim_{z \to \pm \infty} (\Psi^{(n)}(z)) = 0 \text{ for } n=0, 1, 2, \ldots.$$

To geometrically illustrate these relations, Figure A-1 contains the first four derivatives of $\Psi(z)$ as well as $\Psi(z)$ itself in composite form. Clearly, $\Psi(z)$ and all its derivatives of even order are even functions of z, while all the derivatives of odd order are odd functions of z. Furthermore, $\Psi(z)$ is a single valued positive function with maximum value of $\frac{1}{\sqrt{2\pi}}$ at z=0 with points of inflection at z=±1 which approach the abscissa axis asymptotically in both the positive and negative directions. With regard to the first derivative,

$$\Psi^{(1)}(z) = -\Psi(z)H_1(z),$$

the maximum and minimum values are located in the neighborhood of z=-0.9 and z=0.9, respectively, with similar asymptotic tail properties. By comparison, the second derivative possesses a minimum at z=0 and the third derivative has approximate first minima and maxima at z=-0.7 and z=0.7, respectively, while maintaining the required asymptotic tail property. This same geometric behavior is illustrated for the fourth derivative. In this case, a major maximum at z=0 is attained and $\Psi^{(4)}(z)$ crosses the abscissa axis from positive to negative in the neighborhood of z=±0.75.

As a result, these functional characteristics resemble the sine and cosine functions encountered in Fourier series harmonic analysis. This same analogy relates to the functions $\Psi^{(n)}(z)$ and $H_n(z)$ contained in the above noted series. That is, these functional forms characterize a biorthogonal system in an infinite interval. In mathematical terms, such biorthogonal functions satisfy the following conditions:

1) All functions are real and continuous over their plane of definition.
2) No function is identically zero in its plane of definition.
3) Each pair of functions satisfy the integral relation:

$$\int_{-\infty}^{\infty} H_m(z)\Psi^{(n)}(z)dz=0 \quad \text{for } n \neq m.$$

To prove this self-evident relation, note that $\Psi^{(n)}(z) = (-1)^n \Psi(z)H_n(z)$ and $\Psi^{(m)}(z) = (-1)^m \Psi(z)H_m(z)$ so that

$$\int_{-\infty}^{\infty} H_m(z)\Psi^{(n)}(z)dz = (-1)^n \int_{-\infty}^{\infty} H_m(z)H_n(z)\Psi(z)dz$$

$$= (-1)^{n-m} \int_{-\infty}^{\infty} H_n(z)\Psi^{(m)}(z)dz.$$

A-4
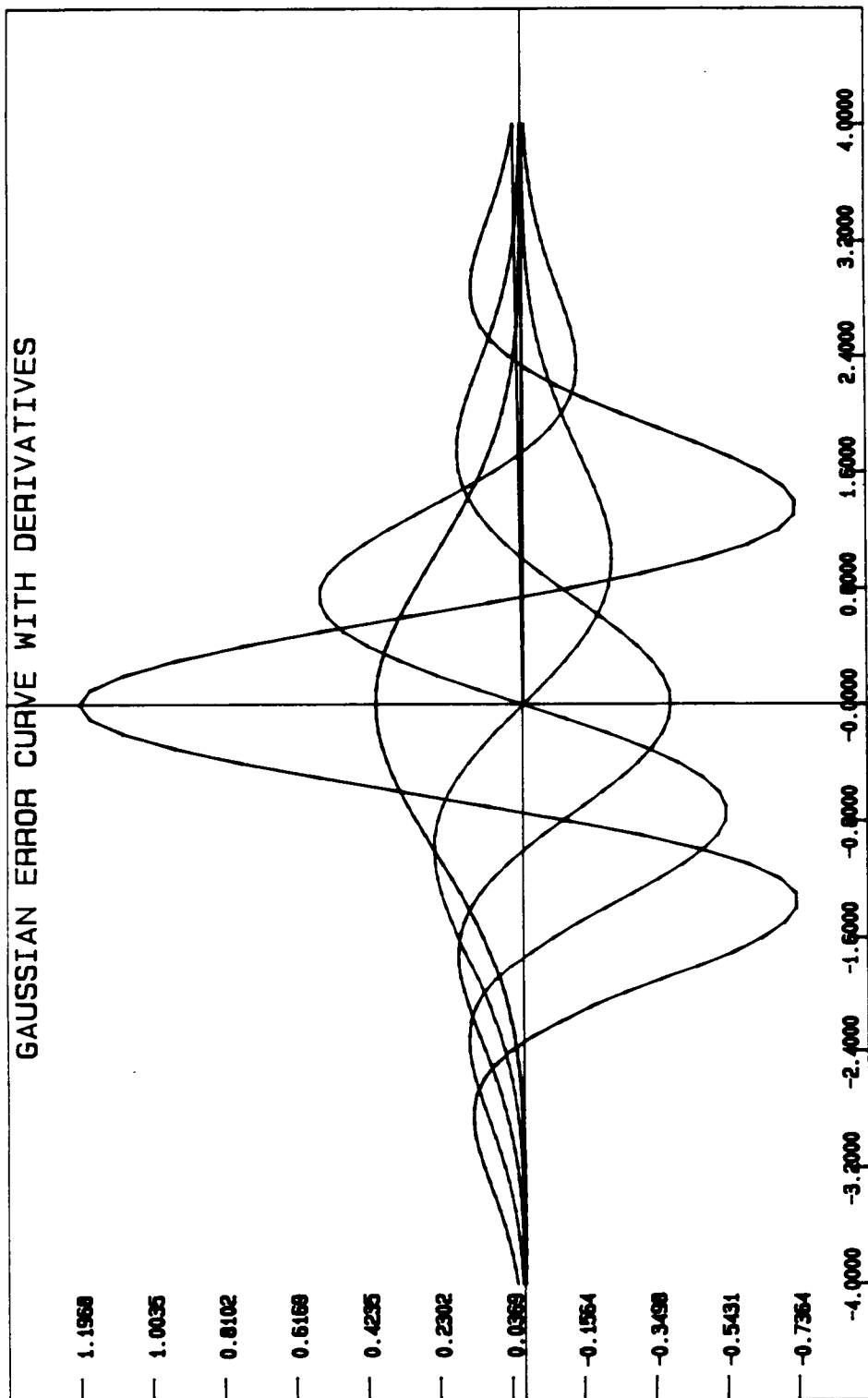
Figure A-1. Composite Gaussian Error Curve with up to Four Derivatives

Since these integral equations hold for all values of m, n = 0, 1, 2, ..., it is only necessary to prove the proposition for n>m.

By the application of integration by parts

$$\int_{-\infty}^{\infty} H_m(z)\Psi^{(n)}(z)\,dz = H_m(z)\Psi^{(n-1)}(z)\Big|_{-\infty}^{\infty} - \int_{-\infty}^{\infty} H_m^{(1)}(z)\Psi^{(n-1)}(z)\,dz$$

where the above notation, $H_m^{(1)}(z)$, denotes the first derivative of $H_m(z)$. Clearly, the first term on the right side of the above equation reduces to 0, since $\Psi^{(n-1)}(z)=0$ for $z=\pm\infty$ and because the order of $H_m(z)$ is lower than $\Psi^{(n)}(z)$. Consequently,

$$\int_{-\infty}^{\infty} H_m(z)\Psi^{(n)}(z)\,dz = -\int_{-\infty}^{\infty} H_m^{(1)}(z)\Psi^{(n-1)}(z)\,dz$$

$$= \int_{-\infty}^{\infty} H_m^{(2)}(z)\Psi^{(n-2)}(z)\,dz$$

$$= -\int_{-\infty}^{\infty} H_m^{(3)}(z)\Psi^{(n-3)}(z)\,dz$$

by three repeated applications of partial integration. If this process is continued m+1 times,

$$\int_{-\infty}^{\infty} H_m(z)\Psi^{(n)}(z)\,dz = (-1)^{m+1}\int_{-\infty}^{\infty} H_m^{(m+1)}(z)\Psi^{(n-m-1)}(z)\,dz.$$

Since $H_m(z)$ is a polynomial of mth degree, its m+1 derivative is equal to zero, which provides the required proof that:

$$\int_{-\infty}^{\infty} H_m(z)\Psi^{(n)}(z)\,dz = 0$$

for all values of m and n where n≠m.

If n=m, exactly the same procedure may be utilized, but the partial integration process is stopped after the nth integration, i.e.,

$$\int_{-\infty}^{\infty} H_n(z)\Psi^{(n)}(z)\,dz = (-1)^n\int_{-\infty}^{\infty} H_n^{(n)}(z)\Psi(z)\,dz$$

after replacing m by n and noting that $\Psi^{(n-n)}(z) = \Psi^{(0)}(z) = \Psi(z)$. However, the nth derivative of the nth degree polynomial, $H_n(z)$, has the constant value n! for all n=0, 1, 2, ...

Therefore,

$$\int_{-\infty}^{\infty} H_n(z)\Psi^{(n)}(z)\,dz = \frac{(-1)^n n!}{\sqrt{2\pi}} \int_{-\infty}^{\infty} e^{\frac{-z^2}{2}}\,dz = (-1)^n n!$$

as a result of the standard normal distribution property

$$\frac{1}{\sqrt{2\pi}}\int_{-\infty}^{\infty} e^{\frac{-z^2}{2}}\,dz=1.$$

A.3 Weighted Square Error Coefficients

Given the biorthogonality property between the functions $\Psi^{(n)}(z)$ and $H_n(z)$, a generalization of the original series introduced by Gram will be derived which minimizes the weighted square error of the approximating probability density function (pdf).

To facilitate this development, it is appropriate to consider a special case of Gram's original series which utilizes the Gaussian distribution as a generating function:

$$\Phi(z) = \sum_{n=0}^{\infty} a_n \Psi^{(n)}(z).$$

To evaluate these coefficients in the least squares sense, consider the following minimization condition:

$$W(z) = \int_{-\infty}^{\infty} \left[\frac{\Phi(z) - \sum_{n=0}^{\infty} a_n \Psi^{(n)}(z)}{\sqrt{\Psi(z)}} \right]^2 dz = \text{Min}.$$

That is to say, the coefficients are determined such that the sum of the squares of the differences between $\Phi(z)$ and the approximating distribution series is a minimum.

To accommodate this solution, note that

$$\sum_{n=0}^{\infty} a_n \Psi^{(n)}(z) = \sum_{n=0}^{\infty} (-1)^n a_n \Psi(z) H_n(z).$$

On the basis of this condition,

$$G(z) = \frac{\sum_{n=0}^{\infty} a_n \Psi^{(n)}(z)}{\sqrt{\Psi(z)}} = \sqrt{\Psi(z)} \sum_{n=0}^{\infty} (-1)^n a_n H_n(z)$$

so that

$$W(z) = \int_{-\infty}^{\infty} \left[\frac{\Phi(z)}{\sqrt{\Psi(z)}} - G(z) \right]^2 dz$$

$$= \int_{-\infty}^{\infty} \left[\frac{[\Phi(z)]^2}{\Psi(z)} - 2\frac{\Phi(z)G(z)}{\sqrt{\Psi(z)}} + [G(z)]^2 \right] dz.$$

To minimize this expression, impose the standard least squares condition that

$$\frac{\partial W(z)}{\partial a_j} = 0 \text{ for } j=0, 1, 2, \ldots;$$

$$0 = \frac{\partial}{\partial a_j} \int_{-\infty}^{\infty} \frac{[\Phi(z)]^2}{\Psi(z)} dz - 2\frac{\partial}{\partial a_j} \int_{-\infty}^{\infty} \frac{\Phi(z)G(z)}{\sqrt{\Psi(z)}} dz + \frac{\partial}{\partial a_j} \int_{-\infty}^{\infty} [G(z)]^2 dz.$$

A-7

Clearly, the first integral on the right-hand side of this equation equals zero, since the integrand is not a function of a_j. Furthermore, taking the partial derivative with respect to a_j under the integral of the second right-hand term,

$$-2 \int_{-\infty}^{\infty} \frac{\Phi(z)}{\sqrt{\Psi(z)}} \frac{\partial G(z)}{\partial a_j} dz = -2 \int_{-\infty}^{\infty} \frac{\Phi(z)}{\sqrt{\Psi(z)}} \frac{\partial}{\partial a_j} \left[\sqrt{\Psi(z)} \sum_{n=0}^{\infty} (-1)^n a_n H_n(z) \right] dz$$

$$= -2 \int_{-\infty}^{\infty} (-1)^j \Phi(z) H_j(z) dz.$$

Similarly, the last term becomes:

$$\int_{-\infty}^{\infty} \frac{\partial}{\partial a_j} [G(z)]^2 dz = \int_{-\infty}^{\infty} 2G(z) \frac{\partial G(z)}{\partial a_j} dz$$

$$= 2 \int_{-\infty}^{\infty} \left[\sqrt{\Psi(z)} \sum_{n=0}^{\infty} (-1)^n a_n H_n(z) \right] \left[(-1)^j \sqrt{\Psi(z)} \ H_j(z) \right] dz$$

$$= 2 \int_{-\infty}^{\infty} \left[\sum_{n=0}^{\infty} (-1)^n a_n \Psi(z) H_n(z) \right] \left[(-1)^j H_j(z) \right] dz$$

$$= 2 \int_{-\infty}^{\infty} \sum_{n=0}^{\infty} (-1)^n a_n \Psi^{(j)}(z) H_n(z) dz$$

$$= 2a_j j!$$

from the biorthogonality property between $\Psi^{(m)}(z)$ and $H_n(z)$. Equating these integral results to zero and solving for a_j, one obtains the following coefficient relationship:

$$a_j = \frac{(-1)^j}{j!} \int_{-\infty}^{\infty} \Phi(z) H_j(z) dz \quad \text{for} \quad j=0, 1, 2, \ldots.$$

Substituting $C_j = \int_{-\infty}^{\infty} \Phi(z) H_j(z) dz$, one obtains the standard GCS approximation form

$$\Phi(z) = \sum_{n=0}^{\infty} (-1)^n \frac{C_n}{n!} \Psi^{(n)}(z).$$

See Korn for an analogous expression (Reference 7).

A.4 Physical Coefficient Interpretation

From a physical coefficient interpretation, two mathematical concerns must be explored and validated. First, one must illustrate that the individual GCS coefficients exist for convergence, and second, that the series has the necessary properties to become a probability density function (pdf).

In theory, the existence of the series coefficients follow if the first two derivatives of $\Phi(z)$ are finite and continuous in the infinite interval and both of these derivatives vanish at $z=\pm\infty$. In more precise mathematical terms, the series converges at all points of continuity if $\Phi(z)$ is of bounded variation. A most interesting discussion of the convergence properties of this general type of series can be found in Cramer (Reference 8).

To illustrate the second concern, consider the approximating GCS:

$$\Phi(z) = \sum_{n=0}^{\infty}(-1)^n \frac{C_n}{n!} \Psi^{(n)}(z)$$

where

$$C_n = \int_{-\infty}^{\infty}\Phi(z)H_n(z)dz.$$

Given these weighted square error minimizing coefficients, the existence of these coefficients and the necessary conditions for a pdf can be unified.

Since it is necessary (but not sufficient) to have

$$\int_{-\infty}^{\infty}\Phi(z)dz = 1$$

for $\Phi(z)$ to be a pdf, it follows that

$$1 = \int_{-\infty}^{\infty}\left[\sum_{n=0}^{\infty}(-1)^n \frac{C_n}{n!} \Psi^{(n)}(z)\right] dz$$

$$= \sum_{n=0}^{\infty}\left[\frac{C_n}{n!} \int_{-\infty}^{\infty} \Psi(z)H_n(z)dz\right]$$

by taking the integral under the sum and then utilizing the identity that $\Psi^{(n)}(z) = (-1)^n\Psi(z)H_n(z)$. This integral sum interchange is valid if and only if the referenced series is convergent. By employing the biorthogonality property between $\Psi^{(n)}(z)$ and $H_m(z)$;

$$\int_{-\infty}^{\infty} \Psi(z)H_n(z)dz = \begin{cases} 0, & \text{for } n=1,\ 2,\ 3,\ \ldots \\ 1, & \text{for } n=0. \end{cases}$$

Since the GCS is convergent, the necessary condition for $\Phi(z)$ to be a pdf imposes the condition that $C_0=1$.

A slightly different type of investigation can be used to access the implications of the remaining coefficients. Here the statistical terminology of moments will be extremely useful. In this terminology, the quantity

$$\mu_k'(b) = E[(z-b)^k]$$

(when it exists) is called the kth moment of the random variable z about b. In integral terms, this statistical expectation operator becomes:

$$E[(z-b)^k] = \int_{-\infty}^{\infty} (z-b)^k f(z)dz,$$

where f(z) is a pdf. Whenever b=0, the expectation $E[z^k]$ is written as

$$\mu_k = \mu_k'(0) = E[z^k] = \int_{-\infty}^{\infty} z^k f(z)dz$$

to denote the kth moment about z=0. In addition, the first moment of a random variable about its expectation $E[(z-\mu_1)]$ is equal to zero. To deduce the implication of this statement, utilize the linearity property of the expectation operator (i.e., $E[ax+by] = aE[x]+bE[y]$) to obtain:

$$E[(z-\mu_1)] = E[z]-E[\mu_1] = 0.$$

The last step of the above reduction results from the fact that the expectation of a constant equals that constant, e.g., $E[\mu_1]=\mu_1$.

In distribution terms μ_1 provides a measure of the distribution centering or expected location. In physical terms μ_1 corresponds to the distribution center of gravity. To illustrate this point, let the z-axis be thought of as a bar with variable density. Under this interpretation the density of the bar at any point is given by f(z) so that $\mu_1 = E[z]$ is the center of gravity of the bar. Therefore, μ_1 may be thought of as a central value of the distribution. Consequently, moments of a distribution about their expected value are called central moments.

Mathematically, the second moment $E[(z-\mu_1)^2]$ is called the variance of the distribution and is often written as σ_z^2. To interpret the physical implications of this terminology in terms of an arbitrary distribution, f(z),

$$\sigma_z^2 = E[(z-\mu_1)^2]$$

$$= \int_{-\infty}^{\infty} (z-\mu_1)^2 f(z)dz$$

$$= \int_{-\infty}^{\infty} [z^2-2z\mu_1+\mu_1^2] f(z)dz$$

$$= \int_{-\infty}^{\infty} z^2 f(z)dz - 2\mu_1 \int_{-\infty}^{\infty} z f(z)dz + \mu_1^2 \int_{-\infty}^{\infty} f(z)dz.$$

Since $\int_{-\infty}^{\infty} f(z)dz=1$ for f(z) to be a pdf, the terms

$$\mu_1 = \int_{-\infty}^{\infty} z f(z)dz \quad \text{and} \quad \mu_2 = \int_{-\infty}^{\infty} z^2 f(z)dz$$

are respectively the first and second moments of $f(z)$. Back substitution into the above expection:

$$\sigma_z^2 = \mu_2 - 2\mu_1^2 + \mu_1^2 = \mu_2 - \mu_1^2.$$

As a result, σ_z^2 is a measure of the spread or dispersion of the distribution about its center of gravity or expected value.

Expressing this central moment notation in specific GCS coefficient terms by equating $\lambda_k = \int_{-\infty}^{\infty}(z-\mu_1)^k \Phi(z)dz$ one obtains the original semi-invariant (moment) equations deduced by Thiele:

$$\lambda_1 = 0$$
$$\lambda_2 = \mu_2 - \mu_1^2$$
$$\lambda_3 = \mu_3 - 3\mu_2\mu_1 + 2\mu_1^3$$
$$\lambda_4 = \mu_4 - 4\mu_3\mu_1 - 3\mu_2^2 + 12\mu_2\mu_1^2 - 6\mu_1^4$$
$$\vdots \qquad\qquad \vdots$$
$$\lambda_N = \mu_N - \sum_{n=0}^{N-2} \binom{N-1}{n} \alpha_{n+1}\, \mu_{N-n-1}$$

where $\alpha_1=\mu_1$, $\alpha_2=\lambda_2$, $\ldots$, $\alpha_{N-1}=\lambda_{N-1}$ for $N>1$. Here, the notation $\binom{N-1}{n}$ denotes the standard binomial coefficient

$$\binom{N-1}{n} = \frac{(N-1)!}{n!(N-n-1)!}.$$

Continuing under the realization that

$$z = H_1(z)$$
$$z^2 = H_2(z)+1$$
$$z^3 = H_3(z)+3z$$
$$z^4 = H_4(z)+6z^2-3$$
$$\vdots \qquad\qquad \vdots$$
$$z^n = H_N(z) - \sum_{n=1}^{[\frac{N}{2}]} (-1)^n \frac{N!}{n!\,2^n (N-2n)!}\, z^{N-2n}$$

for $N>1$ where $\left[\frac{N}{2}\right]$ denotes the greatest integer function, one obtains:

$$\mu_1 = \int_{-\infty}^{\infty} z\Phi(z)dz$$

$$= \int_{-\infty}^{\infty} H_1(z)\Phi(z)dz$$

$$= C_1.$$

This identity was acquired by substituting $z=H_1(z)$ and then using the weighted square error coefficient relationship developed in Section A.3. This same telescoping moment generating procedure produces the following moment coefficient identities:

$$\mu_2 = C_2+1$$
$$\mu_3 = C_3+3C_1$$
$$\mu_4 = C_4+6C_2+3$$
$$\vdots \qquad \vdots$$
$$\mu_N = C_N+ \sum_{n=1}^{\left[\frac{N}{2}\right]} \frac{N!}{n!2^n(N-2n)!} C_{N-2n}$$

for $C_0=1$ and $N>1$. Solving the above identities for the coefficients in terms of the various moments,

$$C_1 = \mu_1$$
$$C_2 = \mu_2-1$$
$$C_3 = \mu_3-3\mu_1$$
$$C_4 = \mu_4-6\mu_2+3$$
$$\vdots \qquad \vdots$$
$$C_N = \mu_N+\sum_{n=1}^{\left[\frac{N}{2}\right]}(-1)^n \frac{N!}{n!2^n(N-2n)!}\mu_{N-2n}$$

for $\mu_0=1$ and $N>1$. However, these relationships are nothing more than the expected value of the referenced order Hermite polynomials, i.e.,

$$C_1 = E[H_1(z)] = E[z] = \mu_1$$

and

$$C_N = E[H_N(z)]$$
$$= E[z^N]+\sum_{n=1}^{\left[\frac{N}{2}\right]}(-1)^n \frac{N!}{n!2^n(N-2n)!} E[z^{N-2n}]$$
$$= \mu_N+\sum_{n=1}^{\left[\frac{N}{2}\right]}(-1)^n \frac{N!}{n!2^n(N-2n)!}\mu_{N-2n}, \quad \text{for } N>1.$$

In summary, it is this observation that makes it convenient to generate the GCS coefficients from a data sample drawn from a parent measurement population. In practice, this "distribution-free" series expansion expresses the required distribution in terms of the derivatives of the standardized normal distribution (with zero mean and unit variance), where the series coefficients are linear combinations of the random data sample moments.

To obtain a physical interpretation of these coefficients, one should note that $C_1=0$ and $C_2=0$ as a result of the transformation

$$z = \frac{x-\mu}{\sigma}$$

from measurement coordinates to z coordinates. Here,

  x = the random measurement sample variate.

  μ = the unbiased measurement mean estimate.

  σ = the parent measurement population standard deviation estimate.

Knowing that the odd order derivatives of $\Psi(z)$ are odd functions and the even order derivatives of $\Psi(z)$ are even functions, the effects of the remaining coefficients may be interpreted. Excluding the first three coefficients of the series, the addition of odd derivative terms (i.e., with their respective C_3, C_5, C_7, ... coefficients) tends to produce asymmetry or skewness relative to the standardized normal distribution. On the other hand, the addition of even derivative coefficient terms tend to flatten the resulting distribution around its mean value.

Technically, the C_3 and C_4 coefficients have statistical names. The C_3 coefficient is called the distribution skewness, while C_4 is the statistical excess or kurtosis. For example, if a unimodal distribution (i.e., with a single maximum) has a longer tail to the right of the central maximum than to the left, the distribution is said to be skewed to the right or to have positive skewness. If the reverse is true, it is said to be skewed to the left or have negative skewness. Thus, distributions which are perfectly symmetrical about their mean (i.e., such as the standardized normal) have zero values for the C_3 coefficient. Consequently, a true normal distribution is characterized by only three series terms, $C_0=1$, $C_1=0$ and $C_2=0$.

APPENDIX B

DISCRETE VERSUS CONTINUOUS FREQUENCY DISTRIBUTIONS

Appendix B
Discrete Versus Continuous Frequency Distributions
Table of Contents

APPENDIX B

DISCRETE VERSUS CONTINUOUS FREQUENCY DISTRIBUTIONS

B.1 Discrete Frequency Concerns

In statistical analysis, attention is often directed toward the enumeration of grouped data which have some particular attribute or characteristic. For instance, measurements falling into a pre-assigned interval are considered to have a common length attribute. A table of the relative proportions of measurements in each class interval specifies the relative frequency distribution or histogram of the inhabitants in each of the discrete class intervals.

The practical difficulties involved in dealing with discrete data arrangements or groupings are that the relative proportion accuracy of grouped data cannot be relied on to produce smooth histogram representations of the true distribution proportions. This is partly due to the discrete interval quantization of finite measurements and the fact that the class measurement proportions are generally considered concentrated at the class interval mid-points. In theory, this latter concern can be relieved by utilized numerical quadrature (integration) to compensate for lumping the entire class interval frequency at its mid-point.

B.2 Numerical Quadrature Frequency Adjustment

As noted above, numerical quadrature can be used to adjust for the concentration of histogram frequencies at their respective class interval mid-points. To explore the ramifications of this adjustment technique, one must first review the continuous distribution properties of moments.

Letting $f(x)$ be a continuous distribution, the integral

$$\int_{-\infty}^{\infty} f(x)\,dx = 1$$

characterizes the probability of the distribution $f(x)$. Consequently, the contribution of any x_j centered within a finite interval of length h in the domain of integration is given by

$$\int_{x_j-\frac{h}{2}}^{x_j+\frac{h}{2}} f(x)\,dx.$$

In contrast, the discrete element of area or probability attained from a finite histogram representation of class interval length h and concentrated frequency f located at x_j is hf.

Extending this elemental notion to distribution moments,

$$\mu_k = \int_{-\infty}^{\infty} x^k f(x)\,dx$$

becomes the kth moment of $f(x)$ in the notation of Appendix A, Section A.4 and its discrete histogram counterpart is

$$M_k = \sum_{j=1}^{N} x_j^k f_j.$$

The basic problem is the development of equivalency approximations between these elemental forms.

Under the supposition that $f(x)$ has close contact with the x-axis at both extremities of the distribution:

1) $\int_{-\infty}^{\infty} x^k f(x) dx$ is finite for all positive values of k.

2) $\underset{x \to \pm\infty}{\text{Lim}} [x^j f^{(k)}(x)] = 0$ for all positive integral values of j and k where $f^{(k)}(x)$ denotes the kth derivative of $f(x)$.

The following relationship can be imposed between these elemental forms:

$$\mu_k = \int_{-\infty}^{\infty} x^k f(x) dx = w \sum_{j=1}^{N} x_j^k f_j.$$

Here, w is the constant of proportionality. Utilizing the basic Newton-Stirling formula of divided differences:

$$f(a+xw) = f(a)+x\Delta f(a)+\tfrac{1}{2}x(x-1)\Delta^2 f(a-w)+\tfrac{1}{6}(x+1)x(x-1)\Delta^3 f(a-w)+$$
$$\tfrac{1}{24}(x+1)x(x-1)(x-2)\Delta^4 f(a-2w)+...$$

and rearranging terms

$$f(a+xw) = f(a)+x[\Delta f(a)-\tfrac{1}{2}\Delta^2 f(a-w)]+$$
$$\tfrac{x^2}{2!}\Delta^2 f(a-w)+\frac{x(x^2-1)}{3!}[\Delta^3 f(a-w)-$$
$$\tfrac{1}{2}\Delta^4 f(a-2w)]+\frac{x^2(x^2-1)}{4!}\Delta^4 f(a-2w)+....$$

Replacing the differences of even order within the brackets by differences of odd order and using the divided difference identities,

$$\Delta^2 f(a-w) = \Delta f(a)-\Delta f(a-w)$$
$$\Delta^4 f(a-2w) = \Delta^3 f(a-w)-\Delta^3 f(a-2w)$$
$$\cdot$$
$$\cdot$$
$$\cdot$$

one obtains:

$$f(a+xw) = f(a)+x\left[\frac{\Delta f(a)+\Delta f(a-w)}{2}\right] + \frac{x^2}{2!}\Delta^2 f(a-w)+$$

$$\frac{x^2(x^2-1)}{3!}\left[\frac{\Delta^3 f(a-w)+\Delta^3 f(a-2w)}{2}\right] + \frac{x^2(x^2-1)}{4!}\Delta^4 f(a-2w)+\ldots$$

Upon substituting x_j for a and n for x and then integrating the above expression from $-\frac{1}{2}$ to $\frac{1}{2}$,

$$\int_{-\frac{1}{2}}^{\frac{1}{2}} f(x_j+nw)\,dn = f(x_j)+\frac{1}{3!2^2}\Delta^2 f(x_j-w) - \frac{17}{3.5!.2^4}\Delta^4 f(x_j-2w)+\ldots$$

In particular, when $f(x)=e^x$,

$$\Delta^{2n} f(x_j-nw) = 2^{2n}e^{x_j}\sinh^{2n}\left(\frac{w}{2}\right).$$

Transforming variables by letting $\theta=\frac{w}{2}$ and dividing throughout by e^x, the above expression reduces to

$$\frac{\sinh\theta}{\theta} = 1 + \frac{1}{3!}\sinh^2\theta - \frac{17}{3.5!}\sinh^4\theta +\ldots$$

Using this result in conjunction with the imposed proportionality constant in the original elemental probability form:

$$\int_{x_j-\frac{h}{2}}^{x_j+\frac{h}{2}} f(x)\,dx = w\sum_{j=1}^{N} x_j f_j.$$

With the added approximation that h=w, one can write

$$\frac{1}{w}\int_{x_j-\frac{w}{2}}^{x_j+\frac{w}{2}} f(x)\,dx = \int_{-\frac{1}{2}}^{\frac{1}{2}} f(x_j+nw)\,dn$$

so that

$$\frac{f_j}{w} = \frac{\sinh\theta}{\theta}\, f(x_j).$$

Substituting this relationship into the histogram moment form

$$\sum_{j=1}^{N} x_{j-k}^k f_j = \sum_{j=1}^{N} x_j^k f(x_j+kw),$$

where the terms on both sides of this equality are the same but counted differently. Therefore,

$$\frac{M_k}{w} = \sum_{j=-1}^{N} f_j\,\frac{\sinh\theta}{\theta} = \sum_{j=-1}^{N} f_j\left[1 + \frac{\theta^2}{3!} + \frac{\theta^4}{5!} + \frac{\theta^6}{7!} +\ldots\right]x_j.$$

At this point, it is convenient to introduce the functional E operator notation (see Reference 9) originally used by the English mathematician, W. P. Sheppard, to express functional operations in ordinary arithmetic terms. It is hypothesized that this correspondence resulted in his name being identified with this discrete moment correction procedure in the statistical literature.

For the current application, the E operator takes advantage of a number of functional laws of correspondence that are interrelated with the central difference operator Δ. To see this, note that

$$M_k = \sum_{j=1}^{N} x_j^k f(x_j + kw) = \sum_{j=1}^{N} x_j^k \left[\Delta^2 E^{-1} \right]^n f_j$$

for some arbitrarily large integer n. To evaluate $\Delta^2 E^{-1}$, let $E = e^{\frac{dw}{dx}}$ and $\Delta = e^{\frac{dw}{dx}} - 1$ so that

$$\frac{1}{4} \Delta^2 E^{-1} = \sinh^2 \left[\frac{1}{2} \frac{dw}{dx} \right].$$

Under this variable substitution, $\theta = \frac{1}{2} wD$ where D denotes the standard differential operator $\frac{d}{dw}$ and

$$M_k = w \sum_{j=1}^{N} f_j \left[1 + \frac{w^2 D^2}{3! \, 2^2} + \frac{w^4 D^4}{5! \, 2^4} + \frac{w^6 D^6}{7! \, 2^6} + \cdots \right] x_j^k$$

$$= w \sum_{j=1}^{N} f_j \left[x_j^k + \frac{w^2}{3! \, 2^2} k(k-1) x_j^{k-2} + \right.$$

$$\left. \frac{w^4}{5! \, 2^4} k(k-1)(k-2)(k-3) x_j^{k-4} + \cdots \right].$$

Substituting the relationship

$$\frac{\mu_k}{w} = \sum_{j=1}^{N} x_j^k f_j$$

into the above expansion

$$M_k = \mu_k + \frac{w^2}{3! \, 2^2} k(k-1) \mu_{k-2} + \frac{w^4}{5! \, 2^4} k(k-1)(k-2)(k-3) \mu_{k-4} + \cdots.$$

Taking k=1, 2, 3, ... N in succession

$$M_1 = \mu_1$$

$$M_2 = \mu_2 + \frac{1}{12} w^2 \mu_0$$

$$M_3 = \mu_3 + \frac{1}{4} w^2 \mu_1$$

$$M_4 = \mu_4 + \frac{1}{2} w^2 \mu_2 + \frac{1}{80} w^4 \mu_0$$

$$M_5 = \mu_5 + \frac{5}{6} w^2 \mu_3 + \frac{1}{16} w^4 \mu_1$$

$$\cdot$$
$$\cdot$$
$$\cdot$$

Solving these equations respectively for the μ_k's in terms of the M_k's and noting that $w=h$ along with the condition that $\mu_0 = \int_{-\infty}^{\infty} f(x) dx = \sum_{j=1}^{N} f_j = M_0 = 1$ by the relative frequency assumption:

$$\mu_0 = M_0$$

$$\mu_1 = M_1$$

$$\mu_2 = M_2 - \frac{h^2}{12} M_0$$

$$\mu_3 = M_3 - \frac{h^2}{4} M_1$$

$$\mu_4 = M_4 - \frac{h^2}{2} M_2 + \frac{7h^4}{240} M_0$$

$$\mu_5 = M_5 - \frac{5h^2}{6} M_3 + \frac{7h^4}{48} M_1$$

$$\cdot$$
$$\cdot$$
$$\cdot$$

$$\mu_N = \sum_{n=0}^{N} \binom{N}{n} (2^{1-n}-1) B_n M_{N-n} h^n$$

where the B_n coefficients are Bernoulli numbers.

APPENDIX C

REFERENCES

APPENDIX C

REFERENCES

1. Hasting, C; "Approximation for Digital Computers," Princeton University Press, Princeton, N. J., 1955.

2. Muller, D. E.; "A Method for Solving Algebraic Equations Using an Automatic Computer," ___, Mathematical Tables and Other Aids to Computations, 10, 1956; pp. 208-215.

3. Todd, J.; "Survey of Numerical Analysis," McGraw-Hill Book Company, Inc., 1962; pp. 258-259.

4. Rudemo, H.; "Empirical Choice of Histogram and Kernel Density Estimators," Scandinavian Journal of Statistics, 9, 1982; pp. 65-78.

5. Gram, J. P.; "Development of Series by Means of the Method of Least Squares (Om Raekkeudviklinger)," Doctorial Thesis, University of Copenhagen, 1879.

6. Thiele, T. N.; "Almindelig Jagttagelseslaere," Copenhagen, 1889 (English Translation, "A General Theory of Observations," C. and E. Layton, London, 1903).

7. Korn, G. A. and Korn, T. M.; "Mathematical Handbook for Scientists and Engineers, McGraw-Hill Book Co., New York, 1968, pp. 19.3-4.

8. Cramer, H.; "Mathematical Methods of Statistics," Princeton University Press; Princeton, N. J., 1945.

9. Cogan, E. J. and Norman, R. Z., "Handbook of Calculus, Difference and Differential Equations," Prentice-Hall, Inc. Englewood Cliffs, N. J., 1958; pp. 9-11.